Stochastic Systems

arXiv: arXiv:0000.0000

CASCADE MARKOV DECISION PROCESSES: THEORY AND APPLICATIONS

By Manish Gupta*,

Harvard School of Engineering and Applied Sciences

This paper considers the optimal control of time varying continuous time Markov chains whose transition rates are themselves Markov processes. In one set of problems the solution of an ordinary differential equation is shown to determine the optimal performance and feedback controls, while some other cases are shown to lead to singular optimal control problems which are more difficult to solve. Solution techniques are demonstrated using examples from finance to behavioral decision making.

1. Introduction. For over five decades the subject of control of Markov processes has enjoyed tremendous successes in areas as diverse as manufacturing, communications, machine learning, population biology, management sciences, clinical systems modelling and even human memory modeling [12]. In a Markov decision process (MDP) the transition rates depend upon controls, which can be chosen appropriately so as to achieve a particular optimization goal. The subject of this paper is to explore a class of MDPs where the transition rates are, in addition, dependent upon the state of another stochastic processes and are thus Markov processes themselves. Our purpose is to describe a broad range of optimal control problems in which these so-called *cascade Markov decision processes* (CMDP) admit explicit solutions [1], as well as problems in which dynamic programming is not applicable at all.

Cascade processes are ideal in modeling games against nature. An epidemic control system where infection rates vary in accordance with uncontrollable factors such as the weather is one such case. They are also applicable in behavioral models of decision making where available choices at each step may be uncertain. For example, a behavioral decision-making problem called the "Cat's Dilemma" first appeared in [7] as an attempt to explain "irrational" choice behavior in humans and animals where observed

*Ph.D Candidate in Applied Mathematics, Harvard School of Engineering and Applied Sciences.

MSC 2010 subject classifications: Primary 60J20, ; secondary 90C40

Keywords and phrases: Markov Decision Processes, Continuous-time Markov Processes

preferences seemingly violate the fundamental assumption of transitivity of utility functions [6],[9]. In this problem, each day the cat needs to choose one among the many types of food presented to it so as to achieve a long-term balanced diet goal. However, the pet owner's daily selection of food combinations presented to the cat is random. The cat's feeding choice forms a controlled Markov chain, but the available foods themselves are contingent on the owner's whim. Another example is found in *dynamically-hedged* portfolio optimization, where dynamic (stochastic) rebalancing of allocated weights can be modeled as a controlled Markov chain. However, what reallocations are possible may depend on the current prices of assets, which are themselves stochastic. Such MDP models have the advantage, for example, of being more realistic than their *continuously-hedged* counterpart, which have traditionally been studied using Gauss/Markov models on augmented state spaces [11], [8]. Other examples where CMDP are applicable include queuing systems where service times depend on the state of another queue and models of resource sharing where one process requires exclusivity and another doesn't (e.g., determining the optimal sync rate for an operating system).

While a cascade Markov process can be equivalently represented on the joint (coupled) state space as a non-cascade, the main purpose of this paper is to investigate solutions on *decomposed* state spaces. The main contributions in doing so include:

- Decoupled matrix differential equations as solutions to a variety of fully observable cascade problems involving optimization of the expectation of a utility functional, which are computationally easier to implement than their non-decoupled counterpart, and require solving of a one-point instead of a two-point boundary value problem.
- Reduction of a partially observable cascade optimal control problem to a lower dimensional non-cascade problem (via a process we call *diagonalization*) that facilitates the use of standard optimization techniques on a reduced state space, thereby circumventing the "curse of dimensionality".
- Simpler analysis, via diagonalization, of a class of problems those that involve optimization of a non-linear function of expectation (such as a fairness or diversity index) and a full solution to a particular example of such *singular* optimal control problems.
- A simple toy model for the dynamically-hedged portfolio optimization problem and solutions that can be easily generalized to computationally feasible algorithms for optimal allocation of large scale portfolios.

In addition to having the advantages of being able to efficiently represent large state space Markov processes by factorization to simpler lower dimensional problems and thus derive computationally simpler solutions, our approach of working decomposed representations is generalizable to multi-factor processes, stochastic automata networks [10], and even quantum Markov chains and controls [5],[3].

The particular framework of Markov decision processes closely follows the assumptions and modeling of [1], which are characterized by finite or denumerably many states with perfect state observations and affine dependence of transition rates on controls. The paper is organized as follows. A mathematical framework is first outlined, more details of which are in Appendix A. We then derive solutions to two classes of optimal control problems. In the first case the cost function is a the expectation of a functional, one that can be solved by dynamic programming requiring solution to a one point boundary value problem. The second class is the case where the cost function can not be written as an expectation, a rather non-standard stochastic control problem but one that arises in applications requiring diversification (entropy) maximization or variance minimization and requires solution to two-point boundary value problems. In many cases the latter is a singular optimal control problem. We will then discuss toy examples in each class of problems: a portfolio optimization problem and animal behavior (decision-making) problem. More examples of portfolio optimization and their cascade solutions appear in the Appendix.

2. Cascade Markov Decision Processes.

2.1. *Markov Decision Process Model.* We use the framework of [1] for continuous-time finite-state (FSCT) Markov processes. We assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and right-continuous stochastic processes adapted to a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in T}$ on this space. An FSCT Markov process x_t that is assumed to take values in $\{e_i\}_{i=1}^n$, the set of $n$standard basis vectors in $\mathbb{R}^n$, has the following sample path (Itô) description: descriptions:

$$dx = \sum_{i=1}^{m} G_i x dN_i \tag{2.1}$$

(2.2)

where $G_i \in \mathbb{G}^n$ are *distinct*[1], $\mathbb{G}^n$ being the space of square $n-$matrices of the form $F_{kl} - F_{ll}$ where F_{ij} is the matrix of all zeros except for one in the $i'th$ row and $j'th$ column, and N_i are Poisson counters with rates λ_i. The resulting *infinitesimal generator* that governs the transition probabilities of the process is $P \in \mathbb{P}^n$, the space of all stochastic $n-$matrices and is given by:

$$P = \sum_{i=1}^{m} G_i \lambda_i$$

In a Markov decision process, the transition rates are allowed to depend on $\mathcal{F}_t-$progressively measurable control processes $u = (u_1, u_2 ... u_p)$ in an affine accordance with[2]:

$$\lambda_i = \lambda_{i0} + \sum_{j=1}^{p} \mu_{ij} u_j$$

so that the infinitesimal generator can be written as:

$$P(u) = \sum_{i=1}^{m} G_i \left(\lambda_{i0} + \sum_{j=1}^{p} \mu_{ij} u_j \right)$$

2.2. *Cascade MDP Model.* We are interested in the case where transition rates of $x_t \in \{e_i\}_{i=1}^n$ are themselves stochastic: specifically, they depend on the state of another Markov process, say, $z_t \in \{e_i\}_{i=1}^r$. We will call such a pair to form a **Cascade Markov chain** (CMC) In general, various levels of interactions between two processes x_t and z_t defines a joint Markov process $y_t = z_t \otimes x_t$ that evolves on the product space $\{e_i\}_{i=1}^n \times \{e_i\}_{i=1}^r$ (see Appendix A) but we are specifically interested in CMCs where sample paths of z_t and x_t have the following have the following Ito description (Proposition A.7, Appendix A):

$$dz = \sum_{i=1}^{s} H_i z dM_i \tag{2.3}$$

$$dx(z) = \sum_{i=1}^{m} G_i(z) x dN_i(z) \tag{2.4}$$

[1]If the $G_i's$ are not distinct, then one can combine the Poisson counters corresponding to identical $G_i's$ to get a set of distinct $G_i's$. For example, $G_1 y dN_1 + G_1 y dN_2$ can be replaced by $G_1 y dN$ where $dN = dN_1 + dN_2$, a Poisson counter with rate equal sum of the rates of the counters N_1, N_2

[2]that is, we assume an affine dependence on controls

where $H_i \in \mathbb{G}^r,\ G_i(z) \in \mathbb{G}^n$ and the rates of Poisson counters M_i and N_i are ν_i and λ_i with λ_i depending on the state of z_t. Thus the infinitesimal generators P and C of x_t and z_t (P depends on z_t and $P(z)$ propagates the *conditional* probabilities of x_t given z) are

$$P(z) = \sum_{i=1}^{m} G_i \lambda_i(z) \tag{2.5}$$

$$C = \sum_{i=1}^{s} H_i \nu_i \tag{2.6}$$

In a **Cascade Markov decision process (CMDP)**, we assume the rates λ_i of counters N_i are allowed to additionally depend on $\mathcal{F}_t-$progressively measurable control processes $u = (u_{1,}u_2...u_p)$ in accordance with [3]

$$\lambda_i(z) = \lambda_{i0}^0 + \lambda_{i0}(z) + \sum_{j=1}^{p} \mu_{ij}(z) u_j$$

so that the conditional probability vector $p(z,u)$ [4] of x_t given z evolves as

$$\dot{p}(z,u) = \sum_{i=1}^{m} G_i \left(\lambda_{i0}^0 + \lambda_{i0}(z) + \sum_{j=1}^{p} \mu_{ij}(z) u_j \right) p(z,u)$$

which will be abbreviated as

$$P(z,u) = A_0 + A(z) + \sum_{j=1}^{p} u_j B_j(z) \tag{2.7}$$

$$\dot{p}(z,u) = P(z,u) p(z,u) \tag{2.8}$$

The CMDP model is completely specified by (A_0, A, B_j).

The Admissible Controls, defining $\mathcal{U}$: The requirements on $P(z,u)$ to be an infinitesimal generator *for each* z put constraints on the matrices A_0, A, B_j and impose admissibility constraints on the controls $u_{j.}$ We will require A_0 and A to be infinitesimal generators themselves (for each t and z) and the B_j to be matrices whose columns sum to zero (for each t and z). We also allow the controls to be dependent on z and x which will define the set of admissible controls $\mathcal{U}$ as the set of measurable functions mapping

[3]Each term is, in additional, a function of time t but for clarity explicit dependence on t will not be specified in notation.

[4]same as above.

the space $\{e_i\}_{i=1}^r \times \{e_i\}_{i=1}^n$ to the space of controls $\mathbb{R}^p$ such that the matrix with j^{th} column

$$f_j = A_0 e_j + A(e_k)e_j + \sum_{i=1}^{p} u_i(e_k, e_j) B_j(e_k)$$

for $j = 1..n,\ k = 1..r$ is an infinitesimal generator. Explicit dependence on t is omitted in notation above for clarity.

2.3. *Examples of CMDP.* Two toy examples of CMDP that will be later discussed are outlined below. Some background on the terminology used in description of portfolio optimization is in Appendix C.1.

2.3.1. *Example 1: A Self-Financing Portfolio Model.* In this toy example on portfolio optimization[5] we will assume that there is one bond and one stock in the portfolio, with the bond price being fixed at 1 and the stock having two possible prices 1 and $-1/3$. Thus the price vector takes values in the set $\{(1,1),(1,-\frac{1}{3})\}$. Assume a portfolio that can shift weights between the two assets with allowable weights W of $(0,2),(-1,-1),(0,-2)$ so that the portfolio has a constant total position (of $\frac{-2}{3}$). Further, we allow only weight adjustments of $+1$ or -1 for each asset, and we further restrict the weight shifts to only those that do not cause a change in net value for any given asset price. The latter condition makes the portfolio *self-financing*.

The resulting process can be modeled as a cascade MDP. Let z_t be the (joint) prices of the two assets with prices $(1,1)$, $(1,\frac{1}{3})$ represented as states e_1, e_2 respectively. Let x_t be the choice of weights with weights $(0,2)$, $(-1,-1)$, $(0,-2)$ represented as states e_1, e_2, e_3 respectively. Transition rates of z_t are determined by some pricing model, whereas the rates of x_t which represent allowable weight shifts are controlled by the portfolio manager. The portfolio value $v(z_t, x_t)$ can be written using its matrix representation, $v(z,x) = z^T V x$, where V is

$$V = \begin{pmatrix} 2 & -\frac{2}{3} \\ -2 & -\frac{2}{3} \\ -2 & \frac{2}{3} \end{pmatrix} \tag{2.9}$$

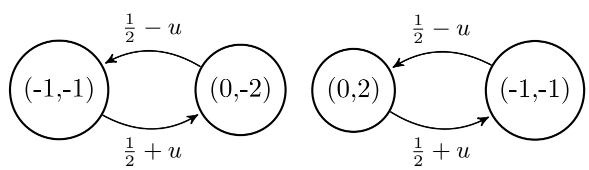
The portfolio manager is able to adjust the rate u of buying stock (which has the effect of simultaneously decreasing or increasing the weight of the bond). The resulting transitions of x_t depend on z_t (see Figure 1(a)) and transition

[5]See Appendix C, Section C.1 for some basic definitions on Portfolio Optimization

matrices $P(z)$ of the weights x_t can be written as $P(z) = A(z)+uB(z)$, where $A(z)$ and $B(z)$are:

$$A(e_1) = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \quad A(e_2) = \frac{1}{2}\begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
$$B(e_1) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \end{pmatrix} \quad B(e_2) = \begin{pmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

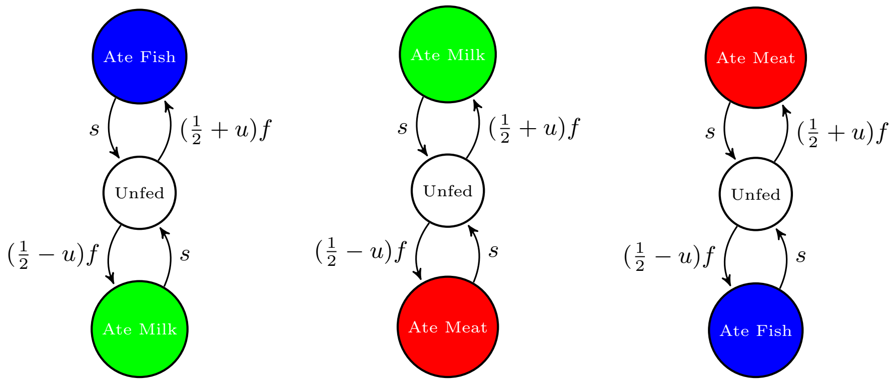
For $P(z)$ to be a proper transition matrix we require admissible controls u needs to satisfy $|u| \leq \frac{1}{2}$. The portfolio manager may choose u in accordance with current values of x_t and z_t so that u is a Markovian feedback controls $u(t, z_t, x_t)$. Note that this model differs from the traditional Merton-like models where only feedback on the total value v_t is allowed. Note that it is the self-financing constraint that leads to the dependence on the current price z_t of the transitions of x, which allows us to model this problem as a cascade.



(a) $z = e_1$ (b) $z = e_2$

FIG 1. *Transition diagram of weight $x(t)$ in the self-financing portfolio for various asset prices $z(t)$ are shown in (a) and (b). States e_1,e_2 of $z(t)$ correspond to price vectors (1,1),(1,-1/3) respectively. Self-transitions are omitted for clarity.*

2.3.2. *Example 2: The Cat's Dilemma Model.* As an example of a cascade MDP, we discuss the cat feeding problem introduced in Section 1. The feeding cat is represented by the process $x(t)$ with four states e_4 =*Unfed*, e_1 =*Ate Meat*, e_2 =*Ate Fish*, e_3 =*Ate Milk*. We assume a constant feeding rate f, and a constant "satisfaction" (digestion) rate s for each food, upon feeding which the cat always returns to the Unfed state. The Markov process $z(t) \in \{e_1, e_2, e_3\}$ represents availability of different combinations of food where e_1, e_2, e_3 denote the combinations $\{$*Fish,Milk*$\}$,$\{$*Meat,Milk*$\}$ and $\{$*Meat,Fish*$\}$ respectively. The food provider is unaffected by the cat's eating rate, and so we can model the process as a cascade with the transition



(a) $z = e_1$ (Fish/Milk) (b) $z = e_2$ (Milk/Meat) (c) $z = e_3$ (Meat/Fish)

FIG 2. *Transition diagram of cat feeding states $x(t)$ in the Cat's Dilemma for various food combinations $z(t)$ are shown in (a), (b) and (c). Self-transitions are omitted for clarity.*

matrix P of x given by (see Figure 2.3.2),

$$P(z,u) = A_0 + A(z) + B(z)u \tag{2.10}$$

where the control $u(z,x) \in [\frac{-1}{2}, \frac{1}{2}]$ represents the cat's choice strategy (extreme values $\pm\frac{1}{2}$ denoting strongest affinity for a particular food in the combination z), and with $A_0, A(z)$ and $B(z)$ given by

$$A_0 = \begin{pmatrix} -s & 0 & 0 & 0 \\ 0 & -s & 0 & 0 \\ 0 & 0 & -s & 0 \\ s & s & s & 0 \end{pmatrix}$$

$$A(e_1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{f}{2} \\ 0 & 0 & 0 & \frac{f}{2} \\ 0 & 0 & 0 & -f \end{pmatrix} \quad A(e_2) = \begin{pmatrix} 0 & 0 & 0 & \frac{f}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{f}{2} \\ 0 & 0 & 0 & -f \end{pmatrix} \quad A(e_3) = \begin{pmatrix} 0 & 0 & 0 & \frac{f}{2} \\ 0 & 0 & 0 & \frac{f}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -f \end{pmatrix}$$

$$B(e_1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & f \\ 0 & 0 & 0 & -f \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad B(e_2) = \begin{pmatrix} 0 & 0 & 0 & -f \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & f \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad B(e_3) = \begin{pmatrix} 0 & 0 & 0 & f \\ 0 & 0 & 0 & -f \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

3. Optimal Control Problem Type I : Expected Utility Maximization.

As alluded to in the introduction, the first category of optimal control problems on cascade MDPs is one where performance measure is the expectation of a functional, and hence linear in the underlying probabilities. We will primarily discuss the fully observable (full feedback), finite time-horizon case and derive a general solution as a matrix differential equation.

3.1. *Problem Definition.* Fix a finite time horizon T on the cascade MDP (z_t, x_t) defined in Section 2.2. and define the cost function η,

$$\eta(u) = \mathbb{E}\int_0^T (z^T(\sigma)\mathbf{L}^T(\sigma)x(\sigma) + \psi(u(\sigma))d\sigma + z^T(T)\Phi^T(T)x(T) \tag{3.1}$$

where c, ϕ are real-valued functions on the space $\mathbb{R}^+ \times \{e_i\}_{i=1}^r \times \{e_i\}_{i=1}^n$, that are represented by the real matrices $\mathbf{L}(t)$ and $\mathbf{\Phi}(t)$ as $c(t,z,x) = z^T\mathbf{L}(t)x$ and $\phi(t,z,x) = z^T\mathbf{\Phi}(t)x$; and ψ a (Borel) measurable function $\mathbb{R}^p \to \mathbb{R}$. If c is bounded the problem of finding the solution to

$$\eta^* = \min_{u\in\mathcal{U}} \eta(u) \tag{3.2}$$

is well-defined and will be subsequently referred to as Problem (**OCP-I**). The corresponding optimal control is given by

$$u^* = \arg\min_{u\in\mathcal{U}} \eta(u) \tag{3.3}$$

3.2. *Solution Using Dynamic Programming Principle.*

THEOREM 3.1. *Let (z_t, x_t) be a cascade MDP as defined in Section 2.2 with C, A_0, A and B_i as defined thereof. Let $T > 0$, and $\mathcal{U}, \psi, \Phi$ and η be as defined in section 3.1. Then there exists a unique solution to the equation (on the space of $n \times r$ matrices)*

$$\dot{K} = -KC - L - A_0^T K - A^T(z)K - \min_{u(z,x)\in\mathcal{U}}(\sum_{i=1}^p u_i z^T K^T B_i(z)x + \psi(u)) \tag{3.4}$$

$$K(T) = \Phi(T) \tag{3.5}$$

*on the interval $[0,T]$, where $A^T(z)K$ denotes the matrix whose j'th column is $A(e_j)K^T e_j^T$ (which can be more explicitly written as $\sum_z A^T(z)Kzz^T$, that is, the matrix representation of the functional $x^T A^T(z)Kz$). Furthermore, if $K(t)$ is the solution to 3.4 then the optimal control problem **OCP-I** defined in (3.2) has the solution*

$$\eta^* = \mathbb{E}z^T(0)K^T(0)x(0) \tag{3.6}$$

$$u^* = \arg\min_{u(z,x)\in\mathcal{U}}(\sum_{i=1}^p z^T K^T u_i B_i(z)x + \psi(u_i)) \tag{3.7}$$

PROOF. With z,x, η as defined above let the minimum return function be $k(t,z,x) = z^T K^T(t)x$,where $K(t)$ is an $n \times r$ matrix, so that $k(0,z(0),x(0) = \eta^*$. Using Ito rule for $z^T K^T x$

$$d(z^T K^T x) = \sum_{i=1}^{s} z^T H_i^T K^T x dM_i + z^T \dot{K}^T x + \sum_{i=1}^{n} z^T K^T G_i x dN_i$$

Since the process $dN_i - (\lambda_{i0}^0 + \lambda_{i0}(z) + \sum_{j=1}^{p} \mu_{ij}(z) u_j)dt$ is a martingale equating the expectation to zero gives

$$\begin{aligned} \mathbb{E}(\sum_{i=1}^{n} z^T K^T G_i x dN_i) &= \mathbb{E}(g(t,x,z,u)dt) \\ \mathbb{E}(\sum_{i=1}^{s} z^T H_i^T K^T x dM_i) &= \mathbb{E}(z^T C^T K^T x) \end{aligned}$$

with $g(t,x,z,u) = z^T K^T A_0 x + z^T K^T A(z) + \sum_{i=1}^{p} z^T K^T u_i B_i(z) x$. Writing $c(t,z,x) + \psi(u) = f(t,z,x,u)$ and $z^T C^T K^T x + g(t,x,z,u) = \xi(t,x,z,u)$, a simple application of the stochastic dynamic programming principle shows that

$$z(t)^T \dot{K}(t)^T x(t) + \min_u (\xi(t,x,z,u) + f(t,z,x,u)) \geq 0$$

The minimum value of 0 is actually achieved by u^*so that the inequality above must be an equality. Introducing notation $A^T(z)K$, we get precisely (3.4). Proof of uniqueness is identical to that in [1] Theorem 1. □

Note that the Bellman equation (3.4) is very similar to that of a single (non cascade) MDP with the additional term $-KC$ representing the backward (adjoint) equation for the process $z(t)$ and the appearance of z in the term for minimization which permits feedback of the optimal control u^* on z in addition to x. The matrix K above is also known as the **Minimum Return Function.** The above solution is a single point boundary value problem instead of two-point. For small KC, the above decouples one column at a time. This form is readily generalizable to multifactor MDPs as well.

COROLLARY 3.2. *(*Quadratic Cost of Control*) Under the hypothesis of the above theorem, if $\psi(u_i) = u_i^2$ then if $u_i(t,z,x) = \frac{-1}{2} z^T(t) K^T(t) B_i(z) x(t)$ lies in the interior of $\mathcal{U}$ then it is the optimal control. Otherwise the optimal control is on the boundary of $\mathcal{U}$. If the former is the case at every $t \in [0,T]$,*

then equation (3.4) defining the optimal solution becomes (where the notation $M^{\cdot 2}$ for a matrix is element-wise squared matrix):

$$\dot{K} = -KC - L - A_0^T K - A^T(z)K + \frac{1}{4}\sum_{i=1}^{p}(B_i^T(z)K)^{\cdot 2}$$

COROLLARY 3.3. (No Cost of Control) *Under the hypothesis of the above theorem, if $\psi(u_i) = 0$ then the optimal control is at the boundary of $\mathcal{U}$. If $\mathcal{U}$ is defined as the set $\{\, -a_i \le |u_i| \le a_i\}$ the optimal control is the bang-bang control $u_i(t,z,x) = -a_i \operatorname{sgn}(z^T K^T(t) B_i(z) x)$ and equation (3.4) defining the optimal solution becomes*

$$\dot{K} = -KC - L - A_0^T K - A^T(z)K + \sum_{i=1}^{p} a_i \left|B_i^T(z)K\right|;$$

3.3. *Solution Using The Maximum Principle.* The stochastic control problem **OCP-I** can be formulated as a deterministic optimization problem (and hence also an open-loop optimization problem) using probability densities permitting the application of variational techniques. While this gives us no particular advantage over the DPP approach in providing a solution to **OCP-I** , understanding this formulation is useful for a broader class of problems.

First we note that for the cascade MDP of Section 2.2 the transition matrices $P(z,u)$ in (2.7) can be written in open-loop form

$$P_i = A_i + \sum_{j=1}^{p} B_{ij} D_{ij} \tag{3.8}$$

where $D_{ij}(u)$ is a diagonal matrix with diagonal $[u_j(e_i,e_1)...u_j(e_i,e_n)]^T$, $A_i = A_0 + A(e_i)$, $B_{ij} = B_j(e_i)$ and $P_i(u) = P(e_j,u)$. Next, we can write evolution of the marginal probabilities $c_i(t) = \Pr\{z(t) = e_i\}$ and joint probabilities $p_{ij}(t) = \Pr\{z(t) = e_i, x(t) = e_j\}$ as the state equations

$$\begin{aligned} \dot{c} &= Cc \\ \dot{p}_i &= P_i p_i + p_i \dot{c}_i / c_i \end{aligned} \tag{3.9}$$

where $p_i(t)$ is the vector $[p_{i1}(t)\, p_{i2}(t)...p_{in}(t)]^T$, $c(t)$ the vector $[c_1(t)...c_r(t)]^T$.

Now we are ready to show the equivalence of the variational approach to the Bellman approach for the problem (**OCP-I**)

THEOREM 3.4. *Let $z \in \{e_i\}_{i=1}^{r}$, $x \in \{e_i\}_{i=1}^{n}$ and $C, A_0, A(z), B_i(z)$ be as defined in Section 2.2 and let the $n-$vectors $p_i(t)$ (for $i = 1..r$) and $r-$vector $c(t)$ satisfy (3.9) with P_i as defined by (3.8) for D_{ij} arbitrary time dependent diagonal $p-$matrices, considered as controls. Then the minimization of*

$$\int_0^T \sum_{i=1}^{r} (e_i^T \mathbf{L}^T p_i + \sum_{j=1}^{p} e^T \psi(D_{ij}) p_i) dt + \sum_{i=1}^{r} e_i^T \Phi^T p_i(T)$$

for $n \times r$ real matrices $\mathbf{L}$ and Φ and (Borel) measurable function $\psi : \mathbb{R}^p \to \mathbb{R}$, subject to the constraints that $P_i \in \mathbb{P}^n$ results in a choice for the k^{th} element of D_{ij} which equals the optimal control $u_j^*(e_i, e_k)$ of Theorem 3.1.

PROOF. Using (3.8) the Hamiltonian H and costate (q, s) for state equations (3.9) for the minimization problem of the theorem become, assuming normality and stationarity of $z(t)$, are

$$\begin{aligned} H &= \sum_{i=1}^{r} q_i^T A_i + \sum_{j=1}^{p} q_i^T B_{ij} D_{ij} p_i + e^T \psi(D_{ij}) p_i + l_i^T p_i \\ \dot{q}_i &= -(A_i + \sum_{j=1}^{p} B_{ij} D_{ij})^T q_i - l_i^T - \sum_{j=1}^{p} \psi(D_{ij}) e \end{aligned} \tag{3.10}$$

where $l_i \equiv e_i^T L^T$, $\phi_i \equiv e_i^T \Phi^T$ and $\psi(D_{ij}) \equiv \psi(u_j(e_i, x))$. Introducing minimization of H with respect to D_{ij} we see that it is achieved by minimizing $\sum_{i=1}^{r} \sum_{j=1}^{p} q_i^T B_{ij} D_{ij} p_i + e^T \psi(D_{ij}) p_i$. Noting that $\psi(D_{ij})$ is also diagonal, simple observation shows that the above is precisely minimized when $\sum_{j=1}^{p} (D_{ij}^T B_{ij}^T q_i + \psi(D_{ij}) e)$ is minimized for each i (as $p_{ik} \geq 0$). The maximum priniciple thus gives the following necessary condition for optimality:

$$\dot{q}_i = -A_i^T q_i - l_i^T - \min_{D_{ij}} (\sum_{j=1}^{p} D_{ij}^T B_{ij}^T q_i + \psi(D_{ij}) e)$$

We note that since stationarity of $z(t)$ was assumed, the above equation exactly corresponds to each column of the Bellman matrix equation (3.4) for K, of Theorem 3.1. (Note that the result is valid for non-stationary $z(t)$ as well and algebraic manipulation shows $(\dot{c}_i / c_i)$ terms to correspond to the $-KC$ term in 3.4). □

REMARK 3.5. *Note that in the variational formulation, linearity of the Hamiltonian in the state variable p for the problem **OCP-I** resulted in complete decoupling of the state and costate equations q_i and p_i thereby permitting an explicit solution identical to that of 3.4. However, if we we restrict*

the set of admissible controls in to allow feedback on the state x but not on the state z in problem ***OCP-I*** *we get a non-trivial variant problem, a partial feedback problem, in which case, one can see that in the analysis in Theorem 3.4 the minimization of $\sum_{i=1}^{r}\sum_{j=1}^{p} q_i^T B_{ij} D_{ij} p_i + e^T \psi(D_{ij}) p_i$, in general, does not lead to a decoupling of the state and costate equations.*

3.4. *Example: A Self-Financing Portfolio.* A toy model of portfolio optimization is discussed as an example or problem **OCP-I**. Appendix B has a short background on portfolio theory and also discusses a variety of **OCP-I** problems on different portfolio models. Consider the problem of maximizing the expected terminal value $v(T)$ of the portfolio for a fixed horizon T for the self-financing portfolio model of Section 2.3.1. With x, z, u, d, V, A, B, D as defined thereof, we wish to maximize the performance measure given by

$$\eta(u, d) = \mathbb{E}(v(T))$$

Using Theorem 3.1 we see the solution to this **OCP-I** problem is obtained by solving the matrix equation with boundary condition $K(T) = -V$

$$\dot{K} = -KC - A^T(z)K + \frac{1}{2}\left|K^T B(z)\right| + \frac{1}{2}\left|K^T D(z)\right| \tag{3.11}$$

with the optimum performance measure and controls (in feedback form) given by

$$\begin{aligned}
\eta^* &= z^T(0)K^T(0)x(0) \\
u^*(t, z, x) &= -\frac{1}{2}\,\mathrm{sgn}(z^T K(t)^T B(z)x) \\
d^*(t, z, x) &= -\frac{1}{2}\,\mathrm{sgn}(z^T K(t)^T D(z)x)
\end{aligned}$$

with $K(t)$ being the solution to (3.11). Some solutions for (3.11) and corresponding optimal controls are plotted for $T = 15$ is shown in Figure 3 for various initial conditions (mixes of the assets in the portfolio initially). Results also show that as $T \to \infty$,the value of η^* approaches a constant value of 1.24 regardless of the initial values $z(0), x(0)$. That is the maximal possible terminal value for the portfolio is 1.24. However, we do not see a steady state constant value for the optimal controls $u^*(z, x)$ and $d^*(z, x)$ and that near the portfolio expiration date, more vigorous buying/selling activity is necessary. If the matrix C were reducible or time-varying in our example, multiple steady-states are possible as $T \to \infty$ and the initial trading activity will be more significant.

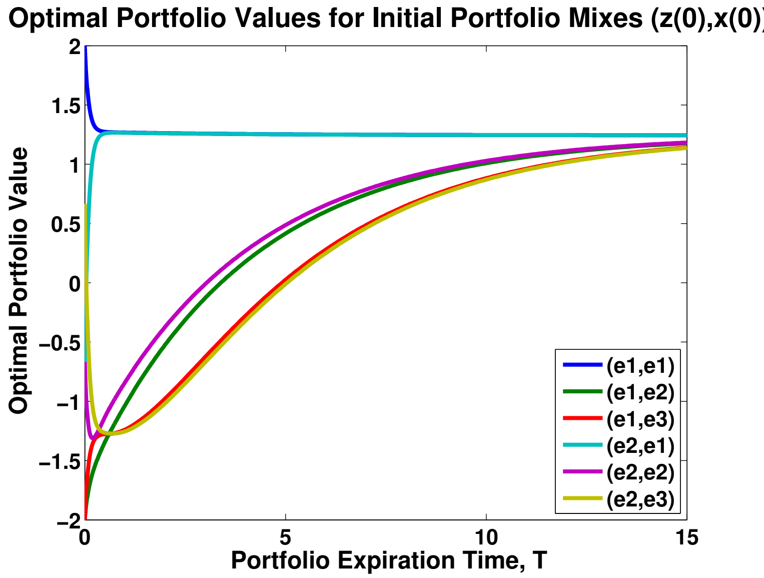


FIG 3. *Minimum Return Function $k(t, z, x)$, for the self-financing portfolio with maximal terminal wealth, shown for various values of z, x specified as (e_i, e_j) vectors.*

Two instances of simulation of application of the above optimal controls are shown in Figures 4 and 5. In the first case we see that one is able to benefit from $x(0)$ being in state e_2 which is the one that corresponds to maximal value of the portfolio, but in which state no trading can take place. We can hold that value and it more than offsets any devaluation due to stock price decline since the stock is more probable to have a higher price than lower. In the second simulation, we are unable to achieve state $x = e_2$, which happens because this state can be attained only in the less probable case of a lower stock price. However, the optimal strategy still trues to maximize the portfolio value by forcing state $x = e_3$ when the price is lower, but since this state is less likely, we need only switch to this sell-out strategy for a small portion of the time. The final value is most sensitive to the final trading activity. The optimal strategy allows us to maximize the portfolio value in all cases, and on the average, gives us the best value.

Our approach of using a cascade model is a more realistic model for portfolio as it is dynamic hedged. Traditional Gauss-Markov models assume continuous hedging which is unrealistic. Our model can be easily extended to include features such as transaction costs, etc. Furthermore, by modeling it as a cascade, we have a computationally scalable solution. The computation time as a function of the dimensionality of the weight sz_t for a decomposed

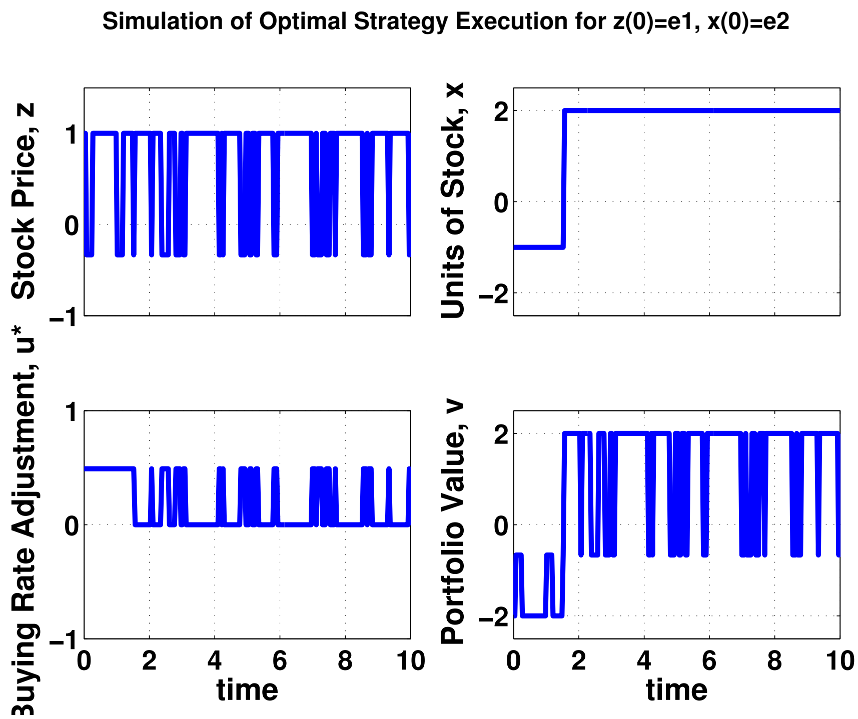


FIG 4. *Simulation 1 of optimal control for self-financing portfolio in 3.4*

.

representation and fully coupled representation (using Bellman equations on the joint process directly) for various expiration times are shown in Figure 6. We see that the solution on a coupled state space grows exponentially with the dimensionality of z_t whereas our solution scales linearly.

4. Optimal Control Problem Type II: Diversification Maximization. The second category of cascade MDP problems are those of optimization of functionals that are non-linear with respect to the probabilities p_{ij}, such as portfolio diversification or fairness of choices in decision making. As alluded to in the introduction, these problems are often singular in the sense that the dynamic programming or maximum principle fail to give a solution, and we will explore this through an example. In general, this class of problems falls in the category where the performance measure to be optimized is a non-linear function of expectation. That is, for a non-linear

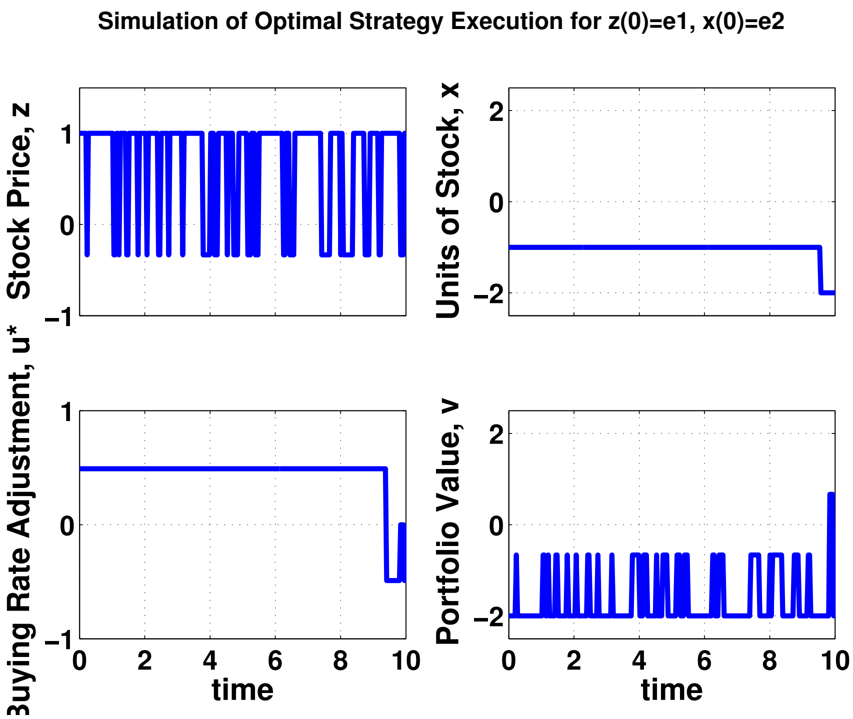
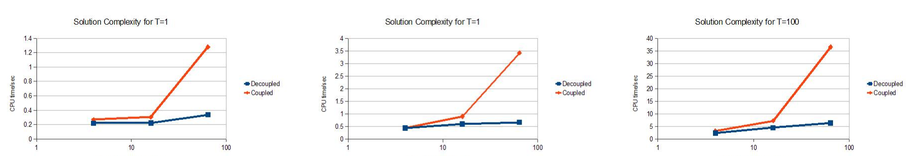


FIG 5. *Simulation 2 of optimal control for self-financing portfolio in 3.4*

.



(a) $T = 1$ (b) $T = 10$ (c) $T = 100$

FIG 6. *CPU time in seconds for asset/bond self-financing toy problem when the number of states of z_t (possible price combinations) increases, for different expiration times $T = 1, 10, 100$. The decoupled solution scales with dimensionality whereas the coupled solution does not.*

f we want to minimize

$$\eta(u) = \int_0^T f(\mathbb{E}(l(t, z_t, x_t, u)))dt \tag{4.1}$$

where $l(.)$ is some loss function. For example, $\eta(u) = \int_0^T (f(x) - \mathbb{E}f(x))^2 \, dt$ minimizes the variance of function f and $\eta(u) = \int_0^T (\mathbb{E}x_1 - \mathbb{E}x_2)^2 \, dt$ specifies adherence to a particular state.

4.1. *Quadratic Problem with No Control Cost.* We formulate a problem that is a particular case of (4.1). Given a cascade $(z_t, x_t, \mathcal{U})$ with model (A, B_i) we define Problem **OCP-II** as the optimal control problem

$$\eta^* = \min_{u \in \mathcal{U}} \lim_{T \to \infty} \frac{1}{T} \int_0^T (p_t^T Q p_t + m^T p_t) dt \tag{4.2}$$

where $Q \geq 0$, m is a vector and p_t is the marginal probability vector of x_t. We note that the stochastic dynamic programming principle is not directly applicable a problem of the form (4.1), and application of variational techniques at best gives us a two-point boundary value problem. Even if we did not have a cascade, the functional of the above form can result in *singular arcs.* To see this heuristically, consider the optimal control problem on a non-cascade defined as

$$\begin{aligned} \dot{p} &= (A + \sum_i u_i B_i) p \\ \eta &= \lim_{T \to \infty} \frac{1}{T} \int_0^T p^T Q p dt \end{aligned} \tag{4.3}$$

with $\mathcal{U} = \{a_i \leq u_i \leq b_i\}$. The costate and Hamiltonian equations for this problem are

$$\begin{aligned} \dot{q} &= -2Qp - A^T q - \left(\sum_i u_i B_i^T \right) q \\ H &= q^T A p + \sum_i q^T u_i B_i p + p^T Q p \end{aligned}$$

so that

$$\frac{\partial H}{\partial u_i} = q^T B_i p$$

If $q^T B_i p = 0$ for any finite time interval, then we have a singular arc so that the Hamiltonian provides no useful information. Characterizing solutions to such singular optimal control problems is notoriously hard. To see how we can get a singular arc in the case above, consider a simplification of (4.3) with $A + uB$ of the form $A + u f_i e_j^T$ for $u \in [a, b]$.For example,

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}^T = (-e_2 + e_3) e_2^T = f_3 e_2^T$$

In steady state, one can show that

$$p(u) = p(0) - \frac{(e_j^T p(0))u \ A^+ f_i}{1 + u(e_j^T A^+ f_i)}$$

where A^+ is the Moore-Penrose inverse of A. If $\eta(u)$ **were** of the form $c^T p$ (the "usual" stochastic control case), then u^* lies on the boundary of $\mathcal{U}$. In the case $\eta(u)$ is of the form $p^T Q p$ (i.e. the non-linear stochastic control case), then it is possible that u^* is in the interior of $\mathcal{U}$. For the class of constant controls, if $u^* \in Int(\mathcal{U})$ then one can show by computation that the corresponding $(p(u), q(u)$ correspond to singular arcs. The above argument heuristically shows why the quadratic control problem OCP-II can be singular.

For a non-cascade, however, the steady state optimal control problem reduces to a non-functional optimization problem, i.e. that of minimizing $p^T(u)Qp(u) + m^T p(u)$. However, for a cascade, η depends on the marginal probabilities of x_t but it is the conditional probabilities $x_t|z_t$ that evolve in accordance with $\dot{p} = Ap$. In general, it is difficult to get an expression for $p(u)$ of the steady state marginal probabilities of x_t but we will below consider a special *diagonalizable* case where $p(u_i)$ satisfy $A_i p = 0$ where p represents the marginal probability vector of x_t.

4.2. *"Cat's Dilemma" Revisited.* In the model presented in Section 2.3.2, the combination of dishes available is random and the cat needs to optimize its selection strategy so as to get a balance of all three dishes. If we assume $s = f = 1$ we note that $\mathbb{E}(x_4) \to \frac{1}{2}$ as $t \to \infty$ regardless of z or u. Hence, the best balance of foods is achieved when each of $\mathbb{E}(x_1), \mathbb{E}(x_2), \mathbb{E}(x_3)$ are as close as possible to $\frac{1}{6}$. Hence the problem can be defined as one of minimizing the performance measure

$$\eta(u) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \|\mathbb{E}(Qx(t,z,u)) - m\|^2 \, dt \tag{4.4}$$

where $\|\|$ is the Euclidean norm on $\mathbb{R}^4$ and Q, m defined as

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} ; \ m = \frac{1}{6} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \tag{4.5}$$

4.3. *A Binary Decision Problem.* The cat's dilemma can be generalized to a class of problems where one needs to make a choice given two possibilities at a time, so as to maximize the diversity of outcomes as a result of

one's choices. If the total number of outcomes is N then binary possibilities are represented by the Markov process $z(t) \in \{e_i\}_{i=1}^r$, $r = \frac{1}{2}N(N-1)$, having generator C where, and the outcomes by the cascade Markov process $x(t) \in \{e_i\}_{i=1}^n$, $n = N+1$, with transition matrix as in (2.10) with r and n dimensional analogs for $A_0, A(z)$ and $B(z)$. The admissibility set $\mathcal{U}$ of controls $u(t,z,x)$ is the set of functions $u : \mathbb{R}^+ \times \{e_i\}_{i=1}^r \times \{e_i\}_{i=1}^n \to [\frac{-1}{2}, \frac{1}{2}]$ such that for each z and t the matrix $P(t,z,u)$ is stochastic. (We can generalize to the situation where, for example, $B(e_{ij}) = (e_i - e_j)^T e_n$ where e_{ij} is choice (i,j) etc.). This cascade model has simpler representations as follows. We will assume $s = f = 1$.

Proposition 4.1. *The model described in Section 4.3 has the following properties.*

1. *(Open loop w.r.t x) For all t, z the dynamics of $x(t)$ do not depend on the controls $u(t,z,x)$ for all $x \neq e_n$*
2. *(Open loop representation w.r.t z) There exist rank one matrices A_j, B_j of the form $f_j e_n^T$ and (open-loop) controls $u_j : \mathbb{R}^+ \to [a,b]$ for $j = 1..r$ such that the transition matrix (2.10) can be written as*

$$P(t, e_j, u) = A_0 + A_j + B_j u_j(t), \text{ for } j = 1..r \tag{4.6}$$

3. *(Triangular Representation) The marginal probabilities $c_j(t) = \Pr\{z(t) = e_j\}$ and $p_k(t) = \Pr\{x(t) = e_k\}$ satisfy the triangular equations*

$$\begin{aligned} \dot{c}(t) &= Cc(t) \\ \dot{p}(t) &= (A_0 + \sum_{j=1}^r c_j(A_j + B_j u_j(t)))p \end{aligned} \tag{4.7}$$

where $p(t) = [p_1(t) \ ...p_n(t)\ ]^T$ and $c(t) = [c_1(t) \ ...c_r(t)\ ]^T$

Proof. Since $B(e_j)$ is of rank one and of the form $f_j e_n^T$ where f_j is a column vector, the dynamics of $x(t)$ depend only the value of control in state $x = e_n$ and z. Thus w.lo.g write $u(t,z,x)$ as $u(t,z)$ instead. Open loop representation (4.6) w.r.t z is made possible by using the parametrization $u_j(t) \equiv u(t, e_j)$ with $B_j = B(e_j)$ and $A_j = A(e_j)$. The triangular representation follows from Corollary B.2 (Appendix B) since $(A_j + u_j B_j)e_k = 0$ for $j = 1..r$, $k = 1..(n-1)$ and that the form of A_0 in (2.10) above above implies that $\Pr(x = e_n)$ is independent of $\Pr(z = e_j)$ for $j = 1..r$. □

The performance measure to maximize diversification of outcomes is (4.4) which can be written using notation introduced in Proposition 4.1 (with $p(t)$

explicitly written as $p(t, u)$ instead),

$$\eta(u) = \lim_{T\to\infty} \frac{1}{T} \int_0^T (Qp(t, u) - m)^T (Qp(t, u) - m) dt \tag{4.8}$$

Two classes of optimal control problems are discussed.

4.4. *Problem 1 : The Steady State Case.* We assume $z(t)$ to be stationary[6]. Using the Given a fixed value T_0 admissibility set $\mathcal{U}_{T_0}$ is restricted to the set of functions $u_j(t), j = 1..r$ that are constant for $t \geq T_0$, as per representation defined in (4.6). With $\eta(u)$ is in (4.8), the optimization problem is

$$\eta^* = \min_{u\in\mathcal{U}_{T_0}} \eta(u), \quad u^* = \arg \min_{u\in\mathcal{U}_{T_0}} \eta(u) \tag{4.9}$$

We will call this problem **OCP-IIS**

THEOREM 4.2. *The solution to the optimization problem* ***OCP-IIS*** *is given by the solution to the quadratic programming problem*

$$\eta^* = \min_u \frac{1}{2} u^T H u + f^T u + k \quad , \textit{subject to} - \frac{1}{2} e \leq I u \leq \frac{1}{2} e$$

where $u \in \mathbb{R}^3, H = \frac{1}{2} A^T A,\ f = A^T b,\ k = b^T b$ *with matrix* A *and vector* b *depending on* $(c_1, c_2 ... c_r)$ *only, and if* u^0 *is the minimizing value for the above, then any function* $u(t)$ *such that* $u(t) = u^0$ *for* $t \geq T_0$ *is an optimal control* u^*.

PROOF. The infinitesimal generator $X(u)$ for $x(t)$ defined in (4.7) is irreducible. Writing the unique time invariant solution to as $p(u)$ a routine calculation shows that

$$p(u) = (ee^T + X^T(u) X(u))^{-1} e \tag{4.10}$$

For $u \in \mathcal{U}_{T_0}$ we can write

$$\begin{aligned} \eta(u) &= \lim_{T\to\infty} \frac{1}{T} (\int_0^{T_0} (Qp(t, u) - m)^T (Qp(t, u) - m)) dt \\ &\quad + \lim_{T\to\infty} \frac{1}{T} \int_{T_0}^T (Qp(u) - m)^T (Qp(u) - m)) dt \\ &= (Qp(u) - m)^T (Qp(u) - m) \end{aligned}$$

[6]If the generator C of $z(t)$ is irreducible then eventually $z(t)$ will attain a time invariant distribution and hence the solution is no different.

Since the first integrand $(Qp(t,u)-m)^T(Qp(t,u)-m))$ is bounded and the second integrand $(Qp(u)-m)^T(Qp(u)-m))$ is independent of t. Using (4.10) write $(Qp(u)-m)^T(Qp(u)-m)=\widetilde{p}^T\widetilde{p}$ where $\widetilde{p}=\frac{1}{2}Au+B$ and A,b are per the statement. Expanding $\widetilde{p}^T\widetilde{p}$ we get the quadratic programming equation. □

CLAIM 4.3. *The quadratic programming equation (Theorem 4.2) has a solution $\eta^*=0$ if and only if the corresponding minimizing value u^0 lies in the interior of the hypercube $[-\frac{1}{2},\frac{1}{2}]^r$*

REMARK 4.4. *Theorem 4.2 can also be proved using explicit computation for the Cat's Dilemma, with $X(u)$ and $p(u)$ given by*

$$X(u) = \begin{bmatrix} -1 & 0 & 0 & c_3\left(u_3+\frac{1}{2}\right)-c_2\left(u_2-\frac{1}{2}\right) \\ 0 & -1 & 0 & c_1\left(u_1+\frac{1}{2}\right)-c_3\left(u_3-\frac{1}{2}\right) \\ 0 & 0 & -1 & c_2\left(u_2+\frac{1}{2}\right)-c_1\left(u_1-\frac{1}{2}\right) \\ 1 & 1 & 1 & -1 \end{bmatrix}$$

$$p(u) = \frac{1}{2}\begin{bmatrix} c_3\left(u_3+\frac{1}{2}\right)-c_2\left(u_2-\frac{1}{2}\right) \\ c_1\left(u_1+\frac{1}{2}\right)-c_3\left(u_3-\frac{1}{2}\right) \\ c_2\left(u_2+\frac{1}{2}\right)-c_1\left(u_1-\frac{1}{2}\right) \end{bmatrix}$$

and A,b thus being computed as

$$A=\begin{bmatrix} 0 & -c_2 & c_3 \\ c_1 & 0 & -c_3 \\ -c_1 & c_2 & 0 \end{bmatrix},\ b=\begin{bmatrix} -\frac{1}{6}+\frac{1}{4}(c_3+c_2) \\ -\frac{1}{6}+\frac{1}{4}(c_1+c_3) \\ -\frac{1}{6}+\frac{1}{4}(c_2+c_1) \end{bmatrix}$$

REMARK 4.5. *We can solve the quadratic programming explicitly. The solutions $u^0\in\mathcal{C}$ where $\mathcal{C}$ is the closed cube $[-\frac{1}{2},\frac{1}{2}]^3$. For the general case of dimensions r and n the results are similar.*

Case 1: When $0<c_j\le\frac{2}{3}$, $j=1..3$.In this case,$\eta^*=0$ and optimal solutions u^0 are given by the lines $u_1=\frac{1}{2c_1}\left(c_2+2c_3u_3+\frac{2}{3}\right), u_2=\frac{1}{2c_2}\left(-c_1+2c_3u_3+\frac{2}{3}\right)$ in the interior of $\mathcal{C}$. **Case 2:** When $c_j\le\frac{2}{3}$ for all j, and $c_j=0$ for some $j,j=1..3$.In this case $\eta^*=0$ and the solutions are given by, for example, in the case $\{\ c_1=0,c_3\le\frac{2}{3}$ and $c_2\le\frac{2}{3}\}$ the set of lines $u_3=-\frac{1}{3c_3}+\frac{1}{2},u_2=\frac{1}{3c_2}-\frac{1}{2}$in the interior of $\mathcal{C}$ but parallel to the faces. **Case 3:** When $\frac{2}{3}<c_j\le 1$ for some j. Since H is singular, several local minima may exist. However, the isolines of global minima are attained along constant values of c_i in the case of $\frac{2}{3}<c_i\le 1$ and the minimal values increase from 0 for $c_i=\frac{2}{3}$ to 0.0408 for $c_i=1$. For example, if $c_2>0$ then

at most two global minima are attained at $u = (0, -\frac{1}{2}, \frac{1}{2})$ or $u = (0, \frac{1}{2}, \frac{1}{2})$ i.e. on the edges of $\mathcal{C}$. If $c_2 = 0$ then the global minimum is attained on the line $u_1 = 0, u_3 = \frac{1}{2}$

4.5. *Problem 2: The Time Varying Case.* Again, we assume $z(t)$ to be stationary. With the admissibility set $\mathcal{U}$ is set of functions $u_j(t), j = 1..r$ such that $u_j(t) \in [-\frac{1}{2}, \frac{1}{2}]$. As per representation defined in (4.6) and with $\eta(u)$ is in (4.8) the problem is

$$\eta^* = \min_{u \in \mathcal{U}} \eta(u), \quad u^* = \arg\min_{u \in \mathcal{U}} \eta(u) \tag{4.11}$$

which we call **OCP-IIT**. In the cases where the steady state optimal control lies in the interior of $\mathcal{U}$, these controls are also optimal within the class of time-varying controls.

PROPOSITION 4.6. *In the cases described in example of Section 4.4 where the solution u^0 to the quadratic programming equation (Theorem 4.2) lies in the interior of the hypercube $[-\frac{1}{2}, \frac{1}{2}]^r$ the solution defined in Proposition 4.2 to **OCP-IIS** for any T_0 is also a solution to **OCP-IIT**.*

PROOF. In the cases of the example of Section 4.4 where the optimal controls are in the interior, optimal performance measure is $\eta^* = 0$. Since the performance measure η defined in (4.4) always satisfies $\eta \geq 0$, thus in these cases a constant control is also optimal within the class of time-varying controls. And this holds for constant controls in the class $\mathcal{U}_{T_0}$ for any finite T_0 (and thus by no means unique). □

4.6. *Singularity Of Optimal Controls.* The problems in Section 4.3 belong to the category of singular control, and an analysis of singularity of optimal solutions presents a slightly more general approach to finding the solution to the time-varying problem (4.11) than the approach above. For this problem, using the representation of Proposition 4.1, the Hamiltonian, state and costate equations can be written as

$$H = (Qp - m)^T (Qp - m) + q^T (A_0 + \sum_{j=1}^{r} c_j A_j + \sum_{j=1}^{r} c_j u_j B_j) p \tag{4.12}$$

$$\dot{p} = (A_0 + \sum_{j=1}^{r} c_j (A_j + B_j u_j(t))) p \tag{4.13}$$

$$\dot{q} = -2(Qp - m) - (A^T + \sum_{j=1}^{r} c_j A_j^T) q - (\sum_{j=1}^{r} c_j u_j B_j^T) q \tag{4.14}$$

However, we see from (4.12) that the costate and state equations are no longer decoupled, and thus trajectories (q, p) that minimize the Hamiltonian can not simply be obtained by solving an equivalent minimization of the individual costate/state equations. In fact, as shown below, we have the case of **singular arcs**, that is, trajectories (solutions) where $q^T B_i p = 0$. Such trajectories fail to give a minimization condition for H with respect to u_i.In such cases, the Maximum Principle at best remains a necessary condition failing to provide the optimal solution. Controls u_i such that the corresponding solutions (p, q) to the state/costate equations form singular arcs will be called **singular controls**.

Proposition 4.7. *For $t > T_0$, the solutions $u^*$to the optimal control problem* ***OCP-IIS*** *that lie in the interior of $\mathcal{U}$ are singular.*

Proof. As $T \to \infty$, $u^*$is a constant control and so p reaches an invariant distribution. Since the optimal trajectory must satisfy the state/costate equation, we see that $\dot{q}$ must be zero as well. Thus, from (4.12) we get by putting $X(u) = \sum_{j=1}^{r}(A_0 + c_j A_j + c_j u_j B_j)$

$$-2(Qp - m) - X^T(u)q = 0$$

Expanding the above for the first $(n-1)$ rows of $X^T(u)q$ we get the equations $q_n - q_i = -2(p_i - \frac{1}{2N})$ for $i = 1..n-1$. These give us the equations $q_i - q_j = 2(p_i - p_j)$ for $i, j = 1..n-1$. The singularity conditions $q^T B_i p = 0$ expand to, by putting in the steady value of $p(u)$, to $q_i - q_j = 0$ for $i, j = 1..n-1$. Since $p_i = p_j = \frac{1}{2N}$ when u^* is in the interior of $\mathcal{U}$ we see that the optimal solutions are singular. □

Thus, in the steady state case, optimal trajectories are singular. We now show that this is also the case for the time-varying case.

Proposition 4.8. *For the problem* ***OCP-IIT****, the value of the Hamiltonian on singular arcs is zero.*

Proof. The state/costate/Hamiltonian are given by (4.12). Without loss of generality, let $p(0) = e_n$. The state equations can be solved explicitly for p_n using $\dot{p}_n = 1 - 2p_n$ to yield $p_n(t) = \frac{1}{2}(1 + e^{-2t})$. Singular arcs satisfy $q^T B_j p = 0$ which expands to $p_n(q_i - q_j) = 0$ for $i, j = 1..(n-1)$ i.e. $q_i = q_j$ for $i, j = 1..n-1$. From $q_i(\infty) = 0$ we get $\dot{q}_i = \dot{q}_j$ or $p_i = p_j$ for $i, j = 1..n-1$ using the costate equation. Using $\sum_{i=1}^{n} p_i = 1$ we get the solution $p_i(t) = \frac{1}{2N}(1 - e^{-2t})$ for $i = 1..(n-1)$. Now plugging these into the costate

equations we can explicitly solve for q_i, $i = 1..n$ for terminal condition $q_i(\infty) = 0$. Omitting details, plugging the solutions into the Hamiltonian, it can be readily seen that $H = 0$. $\square$

COROLLARY 4.9. *The solutions u^* to the optimal control problem (4.11), such that $\lim_{t\to\infty} u^*(t)$ lies in the interior of $\mathcal{U}$, are singular.*

PROOF. In steady state, we see that the optimal trajectories (for which u is in the interior of $\mathcal{U}$) yields $H = 0$ since $(Cp - m)^T(Cp - m) = 0$ and $X(u)p = 0$. From the Maximum Principle, this must be the minimizing value of H and since there is no explicit dependence of H on t this must be the value of H on optimal trajectories at all times. Hence, singular trajectories that satisfy the state/costate equations also minimize the Hamiltonian and so the entire optimal trajectory is singular. $\square$

Now we show that singular solutions are also optimal for the case when optimal controls are in the interior of $\mathcal{U}$.

PROPOSITION 4.10. *For the problem **OCP-IIT** the value of η as defined in (4.8) on singular arcs is zero.*

PROOF. As in the proof of proposition 4.8, on singular arcs, $\frac{\partial H}{\partial u_j} = q^T B_j p = 0$ for $j = 1..r$ give the conditions $q_i = q_j$ for $i, j = 1...(n-1)$. Evaluating $\frac{d}{dt}(\frac{\partial H}{\partial u_j})$ for $j = 1..r$ and setting this to zero (details omitted) yields further the conditions $p_i = p_j$ for $i, j = 1...(n-1)$. Next, evaluating $\frac{d^2}{dt^2}(\frac{\partial H}{\partial u_j})$ for $j = 1..r$ and setting this to zero yields the same equations as in Case 1 and Case 2 of (a generalized version of) the example presented in Section 4.4 .That is, the equations corresponding to $(Qp(u) - m)^T(Qp(u) - m) = 0$ where $p(u)$ is given by (4.10). That is, $\eta = 0$. $\square$

Note that due to the singular nature of the problem, the above analysis does not give us any information about the optimal control u^*. However, we saw from the steady state analysis that a u such that is a constant value satisfying the quadratic programming problem (QPP) (Proposition 4.2) after some finite time is an optimal solution. So if we initially start on a singular trajectory then we remain on it. Otherwise since u is bounded, we can't jump immediately to the singular trajectories and so it will be a bang/bang like control till we transition to an optimal trajectory (though not necessarily constant) control - however, eventually this will become constant. Thus any control that becomes the constant value that is the solution to the QPP in finite time, and one that eventually steers the system onto a singular trajectory is an optimal control.

5. Conclusion. A framework for studying the class of problems where the dynamics of a controllable continuous-time finite-state Markov chain are dependent on an external stochastic process was introduced in this paper and two categories of optimal control problems were discussed. In the "type I" or "expected utility maximization" problems, techniques based upon dynamic programming were used to provide solutions for a general class of problems in the form of a matrix differential equation. This result, proved in Theorem 3.1 using the stochastic dynamic programming, was alternately derived as using the maximum principle, and in the process we were able to see a more general applicability of the variational approach. These solutions were applied to a variety of toy examples in the area of dynamic portfolio optimization. Our factored solutions reduce storage requirements as well as computational complexity significantly. For example, in our representation, a coupled problem with $r = 10, n = 1000$ that would normally require storing a 1000×1000 matrix needs at most ten 100×100 matrices, thereby providing a reduction by a factor of 10. This approach is also generalizable to multi-factor processes, with many interacting Markov chains and with even synchronizing transitions.

Another category of problems, called "type II" or "diversification maximization" problems with performance functionals that are non-linear in underlying state probabilities was discussed in the context of a cat feeding example. It was shown that this problem is singular in the sense that the maximum principle fails to provide an optimal solution, and alternative techniques were explored in the solution of this problem.

Ongoing and future work in this area is focused on general techniques for such singular problems, and extending the class of problems to more complex ones such as multi-cascades (a set of multiple inter-dependent Markov chains), hybrid cascades (for instance, a discrete-state Markov chain with dependencies on continuous-state Gauss-Markov processes) and even decision processes in the context of quantum Markov chains or quantum controls. Computational considerations for large scale versions of the toy portfolio examples presented in this paper will also be investigated.

In this paper only the singular control problem defined in Section 4.3 was analyzed. The general problem of minimizing a performance measure of the form

$$\eta = \int_0^T (\frac{1}{2}p^T Qp + c^T p)dt + \frac{1}{2}p^T(T)S_f p(T) + \phi_f^T p$$

on a cascade MDP where $Q, S_f \geq 0$ needs to be investigated. For the time-invariant case, following the analysis in where it was shown that if a mini-

mizer of $c^T p$ is in the interior of the admissibility set $\mathcal{U}$ then it must define a singular arc, we would like to derive a similar result for the above case. We would further like to derive, for the time-invariant case, sufficient conditions for singular arcs to be optimal (i.e. analog of Proposition 4.10).

Future work in this class of singular problems also involves other techniques such as variable transformations, as in [2], the method of singular perturbations (as in [4]), and numerical methods such as Chebyshev-point collocation techniques.

APPENDIX A: MARKOV PROCESSES ON PRODUCT STATE SPACES

We explore representations of a Markov Process y_t that evolves on the product state space $\{e_i\}_{i=1}^r \times \{e_i\}_{i=1}^n$. The sample path $y(t)$ can be written as the tuple $(z(t), x(t))$ where $z(t) \in \{e_i\}_{i=1}^r$ and $x(t) \in \{e_i\}_{i=1}^n$. The corresponding stochastic processes z_t and $x_t$are the **components** of y_t. The transition matrix for x_t may depend on $z(t)$ and hence describes the propagation of the *conditional probability* distribution $p_{x|z}$: The dynamics of component *marginal* probabilities are not necessarily governed by a single stochastic matrix. Different degrees of coupling between x_t and y_t leads to a possible categorization of the joint Markov Process y_t.

Definition A.1. *A Markov process y_t on the state space $\{e_i\}_{i=1}^r \times \{e_i\}_{i=1}^n$is called **tightly coupled** or **non-decomposable** if there exist states (e_i, e_j) and (e_k, e_l) with $i \neq k$ and $j \neq l$ having non-zero transition probability. If all non-zero transition probabilities are between states of the form (e_i, e_j) to (e_i, e_k), or (e_i, e_j) to (e_l, e_j) then y_t is called **weakly-coupled** or **decomposable**.*

Definition A.2. *A decomposable chain on $\{e_i\}_{i=1}^r \times \{e_i\}_{i=1}^n$ where the transition probability from state (e_i, e_j) to (e_l, e_j) does not depend on j, for all i, l, j where $1 \leq i, l \leq r$ and $1 \leq j \leq n$, is called a **Cascade Markov process**[7].*

Definition A.3. *A cascade Markov process on $\{e_i\}_{i=1}^r \times \{e_i\}_{i=1}^n$ where the transition probability from state (e_i, e_j) to (e_i, e_k) does not depend on i, for alli, j, k where $1 \leq i \leq r$ and $1 \leq j, k \leq n$, is called an **Uncoupled Markov Process**.*

[7]In this paper we mainly focus on Cascade Markov processes, and they are closely related to Markov-modulated Poisson processes (MMPPs) which have vast applications in traffic control, operations research and electronics and communications.

Thus, in a decomposable chain, the jumps in the two component processes are uncorrelated. However, the rates of the counters (and hence transition probabilities) in a component can depend on the state of the another component. In a cascade chain, the rates of the first component (z_t) do not depend on the second component x_t. In an uncoupled chain, the component processes z_t and x_t are completely independent. Decomposable Markov chains have *functional* transition rates, that is, the transition rates are state dependent but do not have any synchronous transitions. Non-decomposable Markov chains exhibit *synchronous transitions*: that is, transitions amongst states of x_t and z_t can occur simultaneously.

A.1. Sample Path and Transition Probability Representations. It will be convenient to represent sample paths $y(t)$ using the Kronecker tensor product $y(t) = z(t) \otimes x(t)$ instead of the tuple $(z(t), x(t))$. The state set $y(t)$ then becomes standard basis for $\mathbb{R}^{r\times n}$. Following the model in (2.1) sample paths $y(t)$ have the Ito representation

$$dy = \sum_{i=1}^{q} G_i y dN_i \tag{A.1}$$

where $G_i \in \mathbb{G}^{rn}$ are distinct. Correspondingly, the infinitesimal generator $P \in \widehat{P}_{rn}$ can be written as $P = \sum_{i=1}^{q} G_i \lambda_i$ where λ_i is the rate of counter N_i.The following results relate decomposability of sample path and transition probability representations to the various levels of couplings defined above.

Proposition A.4. *Let the Markov process y_t be defined on the state space $\{e_i\}_{i=1}^{rn}$ with the Ito representation (A.1). Then for each (distinct) G_i, $i = 1..q$ (see notation defined in Appendix D),*

1. *y_t is a decomposable Markov process if and only if G_i can be written as either $G_i = E_i^r \otimes G_i^n$ or $G_i = G_i^r \otimes E_i^n$.*
2. *If G_i can be written as $G_i = E_i^r \otimes G_i^n$ or $G_i = G_i^r \otimes I_n$ then y_t is a cascade Markov process.*
3. *If G_i can be written as $G_i = I_r \otimes G_i^n$ or $G_i = G_i^r \otimes I_n$ then y_t is an uncoupled Markov process*

Proof. 1. To prove sufficiency, write (A.1) as $dy = \sum_{i=1}^{q_1} (E_i^r \otimes G_i^n)(z \otimes x) dN_i + \sum_{j=q_1+1}^{q} (G_j^r \otimes E_j^n)(z \otimes x) dN_j$. Since $(E_i^r \otimes G_i^n)(z \otimes x) = E_i^r z \otimes G_i^n x$ is a rank one tensor, $zx^T$is a rank one matrix with exactly one non-zero row. Thus jumps in N_i change x but not z. Conversely,

jumps in N_i that change both x and z must have $d(zx^T)$ of rank > 1, i.e. $G_i \neq E_i^r z \otimes G_i^n x$ for any E_i^r and G_i^n.
2. In the decomposable change, transitions that change z but not x correspond to terms such as $(G_j^r \otimes I_n)(z \otimes x)dN_j = (G_j^r z \otimes I_n)dN_j$. Thus the transition G_j^r is driven by N_j only, regardless of x. Since G_j^r are distinct for distinct j, each transition in z is independent of the value of x.
3. Follows by repeating the argument of (2) for the terms $(I_r \otimes G_i^n)(z \otimes x)dN_i$

$\square$

PROPOSITION A.5. *Let the Markov process y_t be defined on the joint state space $\{e_i\}_{i=1}^{rn}$ with infinitesimal generator P. Then, as per notation defined in Appendix D,*

1. *If y_t is decomposable, then P can be written in the form $P = \sum_{i=1}^{p_1} E_i^r \otimes B_i^n + \sum_{i=1}^{p_2} B_i^r \otimes E_i^n$ where B_i^n, B_i^r are matrices such that $\sum_{i=1}^{p_1} B_i^n \in \widehat{P}_n$ and $\sum_{i=1}^{p_2} B_i^r \in \widehat{P}_r$.*
2. *If y_t is a cascade Markov process then P can be written as $P = \sum_{i=1}^{p} E_i^r \otimes B_i^n + C \otimes I_n$ where $C \in \widehat{P}_r$, where B_i^n are matrices such that $\sum_{i=1}^{p_1} B_i^n \in \widehat{P}_n$*
3. *If y_t is an uncoupled Markov process then P can be written as*

$$P = I_r \otimes A + C \otimes I_n \tag{A.2}$$

where $A \in \widehat{P}_n$ and $C \in \widehat{P}_r$.

PROOF. 1. For a decomposable chain from Proposition A.4 we can write (with $q_1 = p_1$ and $p_1 + p_2 = q$) $P = \sum_{i=1}^{p_1}(E_i^r \otimes G_i^n \lambda_i) + \sum_{j=p_1+1}^{q}(G_j^r \lambda_j \otimes E_j^n)$. The result follows from the fact that $\sum_{i=1}^{m} G_i^d \lambda_i \in \widehat{P}_d$ for any integers m and d, and by shifting the summation index in the second sum.
2. Follows from Proposition A.4(2) by setting $C = \sum_{j=p_1+1}^{q} G_j^r \lambda_j$ noting that $C \in \widehat{P}_r$
3. Follows from Proposition A.4(1) as above.

$\square$

The transition matrix representation (A.2) above is not unique to an uncoupled Markov process. In fact, any Markov process y_t on joint state space $\{e_i\}_{i=1}^{rn}$ whose transition matrix P can be written in the form (A.2) is said to be **diagonalizable.** We will shortly see some sufficient conditions

for diagonalizability in the context of MDPs. An important property of diagonalizable Markov processes is that the *marginal* probabilities of the component processes evolve in accordance with stochastic matrices given by the diagonal decomposition, and in fact this condition is also sufficient to guarantee diagonalizability:

PROPOSITION A.6. *Given a diagonalizable Markov process $y_t = z_t \otimes x_t$ whose transition matrix has the diagonal representation (A.2), the marginal probability distributions p_z and p_x of the component processes z_t and x_t evolve in accordance with $\dot{p}_z(t) = Cp_z(t)$ and $\dot{p}_x(t) = Ap_x(t)$ respectively.*

Conversely, given a decomposable Markov process $y_t = z_t \otimes x_t$ such that the marginal probability distributions p_z and p_x of z_t and x_t evolve on $\{e_i\}_{i=1}^r$ and $\{e_i\}_{i=1}^n$ in accordance with $\dot{p}_z(t) = Cp_z(t)$ and $\dot{p}_x(t) = Ap_x(t)$ respectively, where $A \in \widehat{P}_n$ and $C \in \widehat{P}_r$, then y_t is diagonalizable with the representation given by (A.2).

From Propositions A.5 and A.4 we get the following:

PROPOSITION A.7. *Let $y_t = z_t \otimes x_t$ be a Markov process in $r \times n$ states where $z_t \in \{e_i\}_{i=1}^r$ and $x_t \in \{e_i\}_{i=1}^n$. Then sample paths of y_t can be written as*

$$dy_t = (z_t \otimes dx_t) + (dz_t \otimes x_t) + (dz_t \otimes dx_t)$$

If y_t is decomposable, *then the sample paths can be decomposed into*

$$\begin{aligned} z_t \otimes dx_t &= z_t \otimes \sum_{j=1}^{m} G_j(z) x_t dN_j(z_t) \\ dz_t \otimes x_t &= \sum_{i=1}^{s} H_i(z) z_t dM_i(z_t) \otimes x_t \\ dz_t \otimes dx_t &= 0 \end{aligned}$$

where $G_j(z) \in \mathbb{G}^n, H_i \in \mathbb{G}^r$ for each z, x and N_j, M_i are doubly stochastic (Markov modulated) Poisson counters. Furthermore, if y_t is a Cascade MC then we get the following decoupled Ito representation

$$\begin{aligned} dz &= \sum_{i=1}^{s} H_i z dM_i \\ dx &= \sum_{i=1}^{m} G_i(z) x dN_i(z) \end{aligned}$$

REMARK A.8. *If y_t is non-decomposable, the term $dz_t \otimes dx_t$ is non-zero, so we can not write sample paths in decoupled form.*

APPENDIX B: DIAGONALIZABLE MARKOV DECISION PROCESSES

B.0.1. *Properties of Diagonalizable MDPs.* If the MDP is diagonalizable, then some simplifications of the solutions presented above are possible. Once again consider optimal control problem (3.2) except that now the cascade is diagonalizable. Using notation of Section 3.3, the joint probabilities p_i satisfy (assume stationarity of $z(t)$)

$$\dot{p}_i = (A_i + \sum_{j=1}^{p} B_{ij} D_{ij}) p_i$$

From Proposition A.6 and the fact that the marginal probability vector of $x(t)$ is $\sum_{i=1}^{r} p_i$ we must have, for some stochastic matrix $\overline{A}$,

$$\sum_{i=1}^{r} (A_i + \sum_{j=1}^{p} B_{ij} D_{ij}) p_i = \overline{A} \sum_{i=1}^{r} p_i \tag{B.1}$$

Thus we have the following useful lemma:

LEMMA B.1. *Let $z \in \{e_i\}_{i=1}^{r}$, $x \in \{e_i\}_{i=1}^{n}$ and $A(t,z), B_j(t,z), u_j(t,z,x)$, $j = 1..p$, and a cascade MDP on $z \otimes x$ be as defined in Section 2.2. As before, use shorthand $A_i \equiv A(t,e_i)$, $B_{ij} \equiv B_j(t,e_i)$, and $u_j(t,e_i,x)$ as the diagonal matrix D_{ij}.Then the resulting Markov process is diagonalizable if and only if there exists a stochastic matrix $\overline{A}(t)$ such that the joint probabilities written as vectors $\{p_i(t) = [p_{i1}, p_{i2}...p_{in}]^T, i = 1..r\}$ where $p_{ik}(t) = \Pr\{z(t) = e_i, x(t) = e_k\}$ at each t satisfy the equation (B.1), assuming that $z(t)$ is stationary*[8]

COROLLARY B.2. *(**Sufficient Conditions for diagonalizable MDP**).The cascade MDP defined in the hypothesis of Lemma B.1 is diagonalizable if any of following hold:*

1. *$A(t,z)$, $B_j(t,z)$ and $u_j(t,z,x)$ are independent of z, $j = 1, 2..p$. That is, $\mathcal{U}$ is restricted to the set of measurable functions on the space $\mathbb{R}^+ \times \{e_i\}_{i=1}^{n}$ only (i.e. no feedback allowed on state z)*
2. *For each $x \in \{e_k\}_{k=1}^{n}$ and t, the sum $A(t,z) + \sum_{j=1}^{p} u_j(t,z,x) B_j(t,z)$ is independent of z for all admissible controls u_j.*
3. *For each i,k such that the k'th row of $A_i + \sum_{j=1}^{p} B_{ij} D_{ij}$ does not vanish for all t and admissible controls D_{ij}, the marginal probabilities $p_i^Z(t) \equiv \Pr\{z(t) = e_i\}$ and $p_k^X(t) \equiv \Pr\{x(t) = e_k\}$ are uncorrelated, i.e. $p_{ik}(t) = p_i^Z(t) p_k^X(t)$*

[8]Similar equation can be derived for non-stationary $z(t)$ but not needed in this paper

PROOF. The first and second conditions are trivial. For the third, note that if $p_{ik}(t) = p_i^Z(t)p_k^X(t)$ then we can write the $m'th$ row of the left hand side of (B.1) as in fully expanded form, using notation $(M)_{ij}$ for the $(i,j)^{th}$ entry of matrix M

$$\begin{aligned} & \sum_{k=1}^{n}\sum_{i=1}^{r}\sum_{j=1}^{p}(A_i + B_{ij}D_{ij})_{mk}p_{ik} \\ = \ & \sum_{k=1}^{n}\sum_{i=1}^{r}\sum_{j=1}^{p}(A_i + B_{ij}D_{ij})_{mk}p_i^Z p_k^X \\ = \ & \sum_{k=1}^{n}(\sum_{i=1}^{r}p_i^Z\sum_{j=1}^{p}(A_i + B_{ij}D_{ij}))_{mk}\sum_{l=1}^{r}p_{lk} \end{aligned}$$

Setting $\overline{A} = \sum_{i=1}^{r} p_i^Z \sum_{j=1}^{p}(A_i + B_{ij}D_{ij})$which is readily verified to be a stochastic matrix, the result follows from Lemma B.1. □

B.0.2. *Some Problems on Diagonalizable MDPs.* Note that from (B.1) above, a diagonalizable MDP can be rewritten as a partial feedback problem, by possibly introducing matrices $\overline{A}_0(t), \overline{B_i}(t)$ and controls $\overline{u_j}(t,x)$ such that $\overline{A}(t) = \overline{A}_0(t) + \sum_{j=1}^{\overline{p}} \overline{u_j}(t,x)\overline{B_i}(t)$. Thus all optimal control problems on diagonalizable MDPs are in the category of partial feedback problems.

Consider, once again the optimal control problem (3.2) except that now the MDP is diagonalizable. Simplified solutions are available in the following two cases.

THEOREM B.3. *Let $z \in \{e_i\}_{i=1}^r$, $x \in \{e_i\}_{i=1}^n$ and A_0, A,B_i,$T, \mathcal{U}, \psi$,Φ,L, η,be as defined for the cascade MDP on $z \otimes x$ of Theorem 3.1. In addition, let A,B_i and $\mathcal{U}$ satisfy the hypothesis of Corollary B.2.1. Then if the cost functional L or terminal condition Φ do not depend on z, the to the optimal control problem defined in (3.2) has the solution*

$$\begin{aligned} \eta^* &= \mathbb{E}k^T(0)x(0) \\ u^* &= \arg\min_{u(x)\in\mathcal{U}}(\sum_{i=1}^{p} u_i k^T B_i x + \psi(u)) \end{aligned}$$

where k satisfies the vector differential equation

$$\begin{aligned} \dot{k} &= -A^T k - L^T e_1 - \min_{u(x)\in\mathcal{U}}(\sum_{i=1}^{p} u_i k^T B_i x + \psi(u)) \\ k(T) &= \Phi^T e_1 \end{aligned}$$

PROOF. In this case since we have no dependence of $A_{1,}B_j$ or u_j on z and neither that of L or Φ the Bellman equation (3.4) reduces to the single state Bellman equation (See Theorem 1 in [2]) defined on the state space of $x(t)$. Hence we can use a much simplified version of the Bellman equation to find the optimal control. *Note, however, this does not necessarily imply complete independence,* in the sense that the marginal probabilities may still be correlated. $\square$

REMARK B.4. *Note that in view of the introductory remark in Section B.0.2 the condition requiring satisfaction of hypothesis of Corollary B.2.1 is not necessary for a diagonalizable MDP.*

THEOREM B.5. *Let $z \in \{e_i\}_{i=1}^r$, $x \in \{e_i\}_{i=1}^n$ and $A_0, A, B_i, T, \psi, \Phi, L, \eta$, be as defined for the cascade MDP on $z \otimes x$ of Theorem 3.1. Let $\mathcal{U}$ be restricted to the set of measurable functions on the space $\mathbb{R}^+ \times \{e_i\}_{i=1}^n$ (i.e. no feedback allowed on state z), and further let the MDP satisfy the hypothesis of Corollary B.2.3. Using notation $c_i(t) = \Pr\{z(t) = e_i\}, A_i(t) = A(t, e_i), B_{ij}(t) = B_j(t, e_i)$ the optimal control problem defined in (3.2) has the solution*

$$
\begin{aligned}
\eta^* &= c^T(0)\mathbb{E}k^T(0)x(0) \\
u^* &= \arg\min_{u(x)\in\mathcal{U}} (\sum_{i=1}^p u_i k^T(\sum_{i=1}^r c_i B_{ij})x + \psi(u))
\end{aligned}
$$

where k satisfies the vector differential equation

$$
\begin{aligned}
\dot{k} &= -A_0^T k - (\sum_{i=1}^r c_i A_i^T)k - L^T c \\
&\quad - \min_{u(x)\in\mathcal{U}} (\sum_{i=1}^p u_i k^T(\sum_{i=1}^r c_i B_{ij})x + \psi(u)) \\
k(T) &= \Phi^T c
\end{aligned}
$$

PROOF. In this case, if we examine the Hamiltonian in (3.10) we note that in the term to be minimized becomes $\sum_{i=1}^r \sum_{j=1}^p \sum_{k=1}^n p_{ik}(u_{jk} q_i^T B_{ij})$, (assuming no control cost) But since $p_{ik} = p_k c_i$ where p_k and c_i are the marginal probabilities of $x(t) = e_k$ and $z(t) = e_i$ respectively, this otherwise non trivial minimization becomes trivial since we can now interchange the summation order to write this sum as by writing $B_j = .\sum_{i=1}^r c_i B_{ij}$ $\sum_{k=1}^n p_k \sum_{j=1}^p u_{jk}(\sum_{i=1}^r c_i(q_i^T B_j))$ and since $p_k \geq 0$ we achieve minimization by choosing u_{jk} to minimize $(\sum_{i=1}^r c_i(q_i^T B_j))$. This then becomes the condition for the minimum in the costate equation as well, and hence we

have removed dependency of the costate equation on the state p and so we can solve the costate equation (i.e. this becomes the single state Bellman equation). □

APPENDIX C: PORTFOLIO OPTIMIZATION

C.1. Background: Portfolio Value, Wealth and Investment. A portfolio consists of a finite set of assets (such as stocks or bonds), with the *weight process* x_t denoting the vector of amounts (also called allocations or weights) of the assets. The *price process* z_t denotes the vector of market prices of the assets We define the portfolio value $v(t, z, x)$ as the net value of the current asset holdings for weights x and prices z. If $x(t)$ and $z(t)$ take values in finite sets of standard basis vectors, then v can be represented by the matrix $V(t)$ as $v(t, z, x) = z^T V(t)x$. Using the Ito rule, we can write

$$dv = dz^T V^T x + z^T V^T dx$$

In a non self-financing model, depending on the current value of the portfolio, a weight shift will require buying/selling assets using an investment (or consumption, which is the negative of the investment). If $s(t)$ represents the net **investment** into the portfolio up to time t, the incremental investment is the change in the portfolio value due to weight shift. Hence,

$$ds = z^T V^T dx \tag{C.1}$$

Similarly, the **wealth** of the portfolio (i.e. its intrinsic worth) at time t is defined as $w(t) = v(t) - s(t)$. So that the wealth represents the net effect of changes in asset prices, and we can write

$$dw = dz^T V^T x \tag{C.2}$$

C.2. Self-Financing Portfolio Problem. We assume there are two

stocks S_1 and S_2 whose prices each evolve independently on a state space of $\{-1, 1\}$. Assume a portfolio that can shift weights between the two assets with allowable weights W of $(2, 0), (1, 1), (0, 2)$ so that the portfolio has a constant total position (of 2). Further, we allow only weight adjustments of $+1$ or -1 for each asset, and we further restrict the weight shifts to only those that do not cause a change in net value for any given asset price. The latter condition makes the portfolio self-financing.

The resulting process can be modeled as a cascade MDP. Let z_t be the (joint) prices of the two assets with prices $(-1, -1)$, $(-1, 1)$, $(1, -1)$, $(1, 1)$

represented as states e_1, e_2, e_3, e_4 respectively. Let x_t be the choice of weights with weights $(0, 2)$, $(1, 1)$, $(2, 0)$ represented as states e_1, e_2, e_3 respectively. Transition rates of z_t are determined by some pricing model, whereas the rates of x_t which represent allowable weight shifts are controlled by the portfolio manager. The portfolio value $v(z_t, x_t)$ can be written using its matrix representation, $v(z, x) = z^T V x$, where V is

$$V = \begin{pmatrix} -2 & 2 & 0 & 2 \\ -2 & 2 & -2 & 2 \\ -2 & 0 & -2 & 2 \end{pmatrix} \tag{C.3}$$

The portfolio manager is able to adjust the rate of increasing the first weight by an amount u and, independently that of decreasing the first weight by an amount d (which has the effect of simultaneously decreasing or increasing the weight of the second asset). The resulting transitions of x_t depend on z_t (see Figure C.2) and transition matrices $P(z)$ of the weights x_t can be written as $P(z) = A(z) + uB(z) + dD(z)$, where $A(z), B(z), D(z)$ are:

$$A(e_1) = \tfrac{1}{2}\begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix} \quad A(e_2) = \tfrac{1}{2}\begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad A(e_3) = \tfrac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \quad A(e_4) = \tfrac{1}{2}\begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix}$$

$$B(e_1) = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad B(e_2) = \begin{pmatrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad B(e_3) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad B(e_4) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$D(e_1) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix} \quad D(e_2) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad D(e_3) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix} \quad D(e_4) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

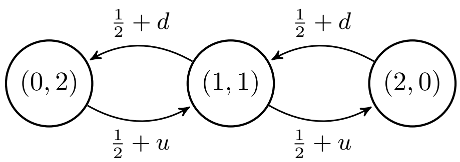
For $P(z)$ to be a proper transition matrix we require admissible controls u, d to satisfy $|u|, |d| \leq \frac{1}{2}$. The portfolio manager may choose u, d in accordance with current values of x_t and z_t so that u, d are Markovian feedback controls $u(t, z_t, x_t)$ and $d(t, z_t, x_t)$. Note that this model differs from the traditional Merton-like models where only feedback on the total value v_t is allowed. Note that it is the self-financing constraint that leads to the dependence on the current price z_t of the transitions of x which allows us to model this problem as a cascade.

Consider the problem of maximizing the expected terminal value $v(T)$ of the portfolio for a fixed horizon T for the above self-financing portfolio model 2.3.1. With x, z, u, d, V, A, B, D as defined thereof, we wish to maximize the performance measure given by

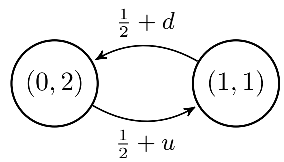
$$\eta(u, d) = \mathbb{E}(v(T))$$

Using Theorem 3.1 we see the solution to this **OCP-I** problem is obtained by solving the matrix equation with boundary condition $K(T) = -V$

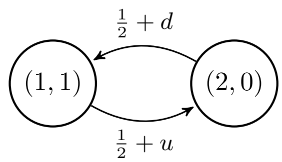
$$\dot{K} = -KC - A^T(z)K + \frac{1}{2}\left|K^T B(z)\right| + \frac{1}{2}\left|K^T D(z)\right| \tag{C.4}$$



(a) $z = e_1$ or $z = e_4$



(b) $z = e_2$



(c) $z = e_3$

FIG 7. *Transition diagram of weight $x(t)$ in the self-financing portfolio for various asset prices $z(t)$ are shown in (a), (b) and (c). States e_1,e_2,e_3,e_4 of $z(t)$ correspond to price vectors (-1,-1),(-1,1),(1,-1),(1,1) respectively. Self-transitions are omitted for clarity.*

with the optimum performance measure and controls (in feedback form) given by

$$\begin{aligned}
\eta^* &= z^T(0)K^T(0)x(0) \\
u^*(t,z,x) &= -\frac{1}{2}\,\mathrm{sgn}(z^T K(t)^T B(z)x) \\
d^*(t,z,x) &= -\frac{1}{2}\,\mathrm{sgn}(z^T K(t)^T D(z)x)
\end{aligned}$$

with $K(t)$ being the solution to (C.4). Some solutions for (C.4) and corresponding optimal controls are plotted for $T = 1, 15$ in Figure 8 for various initial conditions (mixes of the assets in the portfolio initially). Results also show that as $T \to \infty$,the value of η^* approaches a constant value of 0.4725 regardless of the initial values $z(0), x(0)$. That is the maximal possible terminal value for the portfolio is 0.4725. However, we do not see a steady state constant value for the optimal controls $u^*(z,x)$ and $d^*(z,x)$ and that near the portfolio expiration date, more vigorous buying/selling activity is necessary. If the matrix C were reducible or time-varying in our example, multiple steady-states are possible as $T \to \infty$ and the initial trading activity will be more significant.

C.3. An Investment-Consumption Portfolio Problem. An alternate model for portfolio allocation than discussed in the self-financing Portfolio example (Section) is presented as a **OCP-I** problem in this section.

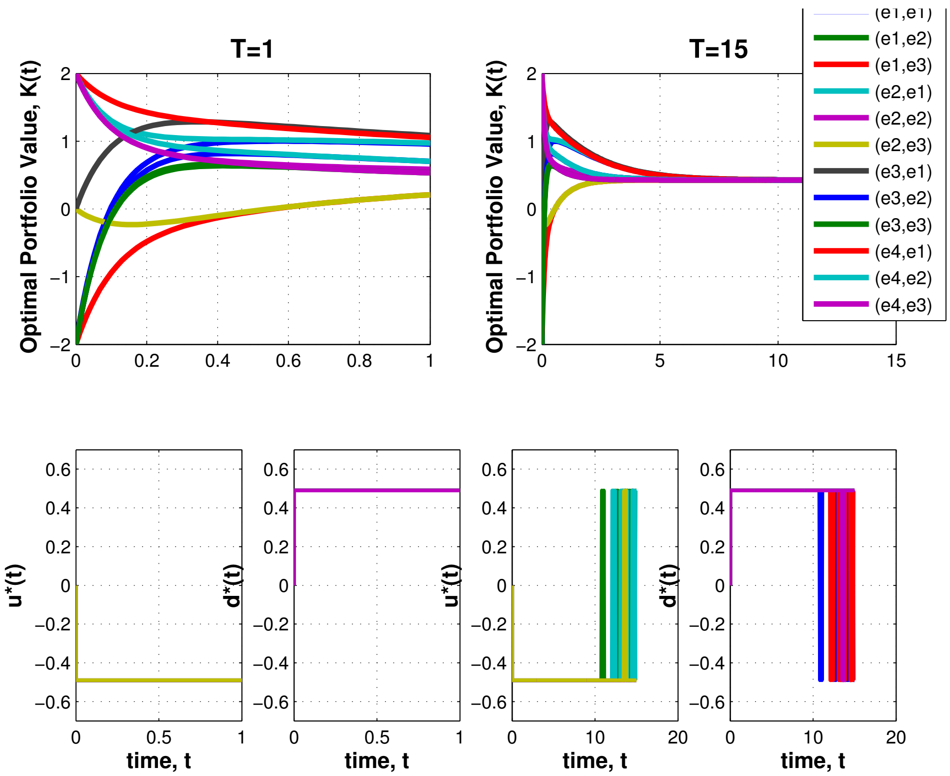


Fig 8. *Solution to problem* C.2. *Minimum Return Function* $k(t, z, x)$, *optimal up controls* $u^*(t, z, x)$ *and down controls* $d^*(t, z, x)$ *for the self-financing portfolio with maximal terminal wealth. Figures(a) and (b) are for* $T = 1$ *and* $T = 15$ *respectively. Various* (z, x) *values are represented by the vectors* (e_i, e_j).

If we do not restrict the weight adjustments in the model of Section 2.3.1 to cases which keep the value a constant, (i.e. we allow only weight adjustments of $+1$ or -1 for each asset, regardless of the current portfolio value) we get a non self-financing portfolio. The difference in the portfolio value as a result of weight shift must be the result of an equivalent investment or consumption. Once again, modeling this is as a cascade with z_t and x_t as in Section 2.3.1, the portfolio value matrix (C.3) is replaced by

$$V = \begin{pmatrix} -2 & 2 & -2 & 2 \\ -2 & 0 & 0 & 2 \\ -2 & 2 & 2 & 2 \end{pmatrix} \tag{C.5}$$

As before, the portfolio manager can control the up and down rates u, d resulting in the transitions of x_t (See Figure) described by the matrices

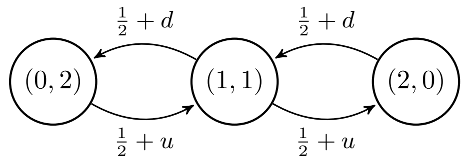


FIG 9. *Transition diagram of weights $x(t)$ for controls u,d in the investment/consumption portfolio. Self-transitions are omitted for clarity.*

$P(z) = A(z) + uB(z) + dD(z)$ with

$$A(z) = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

$$B(z) = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad D(z) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

and admissibility condition $|u|, |d| \leq \frac{1}{2}$. Note in this cascade model the matrices A, B, D do not depend on z but we will see next that the performance measure does depend on z.

C.3.1. *Problem 1: Minimal Investment.* We wish to minimize the total amount of investment up to a fixed horizon T. We can write a performance measure $\eta(u, d)$ that represents the net investment into the portfolio up to time T as

$$\eta(u, d) = \mathbb{E}(s(T))$$

where $s(t)$ is the *investment* process (See Appendix C.1). Using (C.1)

$$\mathbb{E}(ds(t)) = \mathbb{E}(z^T V^T dx) = \mathbb{E}(z^T V^T (A + uB + dD)x)dt$$

Writing the matrix $\Phi(u, d) = V^T(A + uB + dD)$

$$\eta(u, d) = \mathbb{E} \int_0^T z^T(t)\Phi(u, d)x(t)dt$$

The goal then is to choose u, d so as to minimize $\eta(u, d)$ subject to $u, d \in \mathcal{U}$ where the admissibility set $\mathcal{U}$ is the set of past measurable functions $u(z, x)$ such that $|u(z, x)| \leq \frac{1}{2}$ for each z, x. Using Theorem 3.1 we see the solution

to this **OCP-I** problem is obtained by solving the matrix equation with boundary condition $K(T) = 0$

$$\dot{K} = -KC - A^T(K+V) + \frac{1}{2}\left|B^T(K+V)\right| + \frac{1}{2}\left|D^T(K+V)\right| \tag{C.6}$$

(where the notation $|M|$ for a matrix M above represents the element-by-element absolute value of a matrix) with the optimal performance measure and controls (in feedback form) given by

$$\begin{aligned} \eta^* &= z^T(0)K(0)x(0) \\ u^*(t,z,x) &= -\frac{1}{2}\,\mathrm{sgn}(z^T(K(t)+V)^T Bx) \\ d^*(t,z,x) &= -\frac{1}{2}\,\mathrm{sgn}(z^T(K(t)+V)^T Dx) \end{aligned}$$

where $K(t)$ is the solution to (C.6). Some solutions to (C.6) and corresponding optimal controls are plotted for $T = 1, 10$ in Figure 10(a) and 10(b). Results also show that as $T \to \infty$,the value of η^*/T approaches a constant value of -0.535 regardless of the initial values $z(0), x(0)$ and in this case we see that the optimal controls $u^*(z,x)$ and $d^*(z,x)$ expressed in matrix form ($u^*(z,x)$ written as $z^T u^* x$ etc.)

$$u^* = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix};\ d^* = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

(the values of $u^*(z, e_3)$ and $d^*(z, e_1)$ are immaterial as they do not impact the dynamics). This means that one can expect a constant cash flow of 0.535 by the above strategy, and that this value is maximal. Note also that the optimal controls *do* depend on z and so the resulting weight and asset probabilities are not independent.

C.3.2. *Problem 2 : Maximal Terminal Wealth.* In this case the performance measure that needs to be maximized is given by

$$\eta(u,d) = \mathbb{E}(w(T)) = \mathbb{E}\int_0^T z^T(t)C^TV^Tx(t)dt$$

where $w(t)$ is the *wealth* process (Appendix C.1) for $u, d \in \mathcal{U}$ as above. Again, from Theorem 3.1 the solution to this **OCP-I** problem is obtained by solving the matrix equation with boundary condition $K(T) = 0$

$$\dot{K} = -(K-V)C - A^TK + \frac{1}{2}\left|B^TK\right| + \frac{1}{2}\left|D^TK\right| \tag{C.7}$$

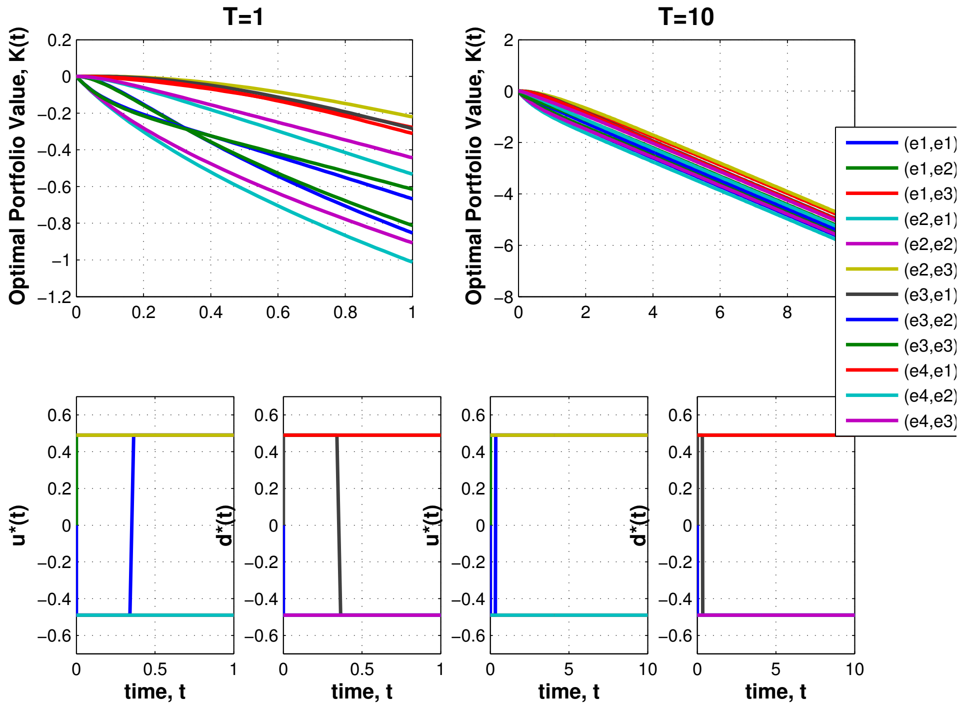


FIG 10. *Solution to problem C.3.1. Minimum Return Function $k(t,z,x)$, optimal up controls $u^*(t,z,x)$ and down controls $d^*(t,z,x)$ for the self-financing portfolio with maximal terminal wealth. Figures(a) and (b) are for $T=1$ and $T=10$ respectively. Various (z,x) values are represented by the vectors (e_i,e_j).*

whose solution $K(t)$ gives the optimal performance measure and controls as:

$$\begin{aligned}\eta^* &= z^T(0)K(0)x(0)\\ u^*(t,z,x) &= -\frac{1}{2}\,\mathrm{sgn}(z^TK^T(t)Bx)\\ d^*(t,z,x) &= -\frac{1}{2}\,\mathrm{sgn}(z^TK^T(t)Dx)\end{aligned}$$

Some numerical results for the above problem with V as in (C.5) are plotted for $T=1,10$ in Figure 11 (a) and 11 (b). Results also show that as $T\to\infty$,the value of η^*/T approaches a constant value of -0.533 regardless of the initial values $z(0),x(0)$ and in this case we see that the optimal controls $u^*(z,x)$ and $d^*(z,x)$ expressed in matrix form ($u^*(z,x)$ written as z^Tu^*x

etc.) are

$$u^* = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix} ; \; d^* = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

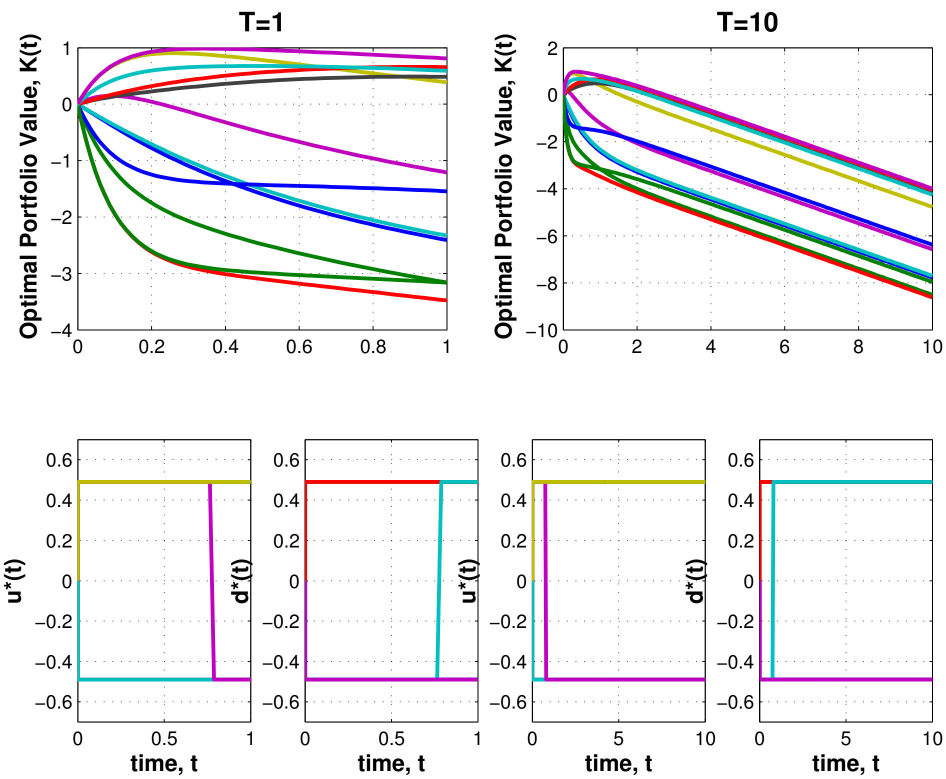


FIG 11. *Solution to problem C.3.2. Minimum Return Function $k(t,z,x)$, optimal up controls $u^*(t,z,x)$ and down controls $d^*(t,z,x)$ for the self-financing portfolio with maximal terminal wealth. Figures(a) and (b) are for $T = 1$ and $T = 10$ respectively. Various (z,x) values are represented by the vectors (e_i, e_j).*

C.3.3. *Problem 3 - Minimal Investment with Partial Feedback.* In the investment/consumption model, the control matrices A, B, D do not depend on z. As a result one may be tempted to think that a partial feedback optimization problem, i.e. where the controls are allowed to depend on x but not z would give the same optimal performance. However, one sees from Theorem 3.1 the solution to the minimal investment case is obtained by

solving the matrix equation subject to $K(T) = 0$

$$\begin{aligned} \dot{p}_z &= Cp_z;\; p_z(0) = \mathbb{E}z(0) \\ \dot{K} &= -KC - A^T(K+V) + \frac{1}{2}\left|B^T(K+V)(\mathbf{e}_r p_z^T)\right| \\ &\quad + \frac{1}{2}\left|D^T(K+V)(\mathbf{e}_r p_z^T)\right| \end{aligned} \tag{C.8}$$

where p_z is the probability vector for z. And the optimal performance and controls are given by

$$\begin{aligned} \eta^* &= p_z^T(0)K^T(0)x(0) \\ u^*(t,x) &= -\frac{1}{2}\operatorname{sgn}((\mathbf{e}_r p_z^T(t))(K(t)+V)^T Bx) \\ d^*(t,x) &= -\frac{1}{2}((\mathbf{e}_r p_z^T(t))(K(t)+V)^T Dx) \end{aligned}$$

where $K(t), p_z(t)$ are solutions to C.8. The best performance in this case is worse than that in the full feedback case, as indeed shown by numerical simulation as in 12(a),(b) for $T = 1, 10$. Comparing with the respective minimum return functions of the full feedback case, the steady state case maximal cash flow rate is only 0.22 compared to 0.533.

C.4. Some Variations on Portfolio Problems. Some variants of the examples presented here and in Section 3.4 include the following:

C.4.1. *Utility Functions.and Discounting.* In traditional portfolio optimization problems, one minimizes $\mathbb{E}(U(s(T)))$o maximizes $\mathbb{E}(U(w(T)))$ where $U(.)$ is a non-decreasing and concave function, called the *utility function*. In the above examples, for simplicity of demonstration of the MDP techniques, we assumed $U(C) = C$. Utility functions are chosen based upon risk preferences of agents and the financial environment, and some standard ones include the $U(C) = \frac{C^\gamma}{\gamma}$ (with $\gamma < 1$) or $U(C) = \log C$. Furthermore, one may wish to optimize the *discounted* value i.e $\mathbb{E}\int_0^T e^{-\alpha t}U(w(t))dt$ for some $\alpha > 0$ instead. The solutions to optimization problems of minimum investment and maximum wealth in these cases are identical to (C.6) and (C.7) with V replaced by $e^{-\alpha t}U(V)$.

C.4.2. *Value Payoff Functions.* The particular model we chose led to a value payoff V as in (C.5) though the problems presented above are completely generic with respect to V in that any other value of V would work as well. In that case we will have different mappings of the states e_1, e_2, e_3

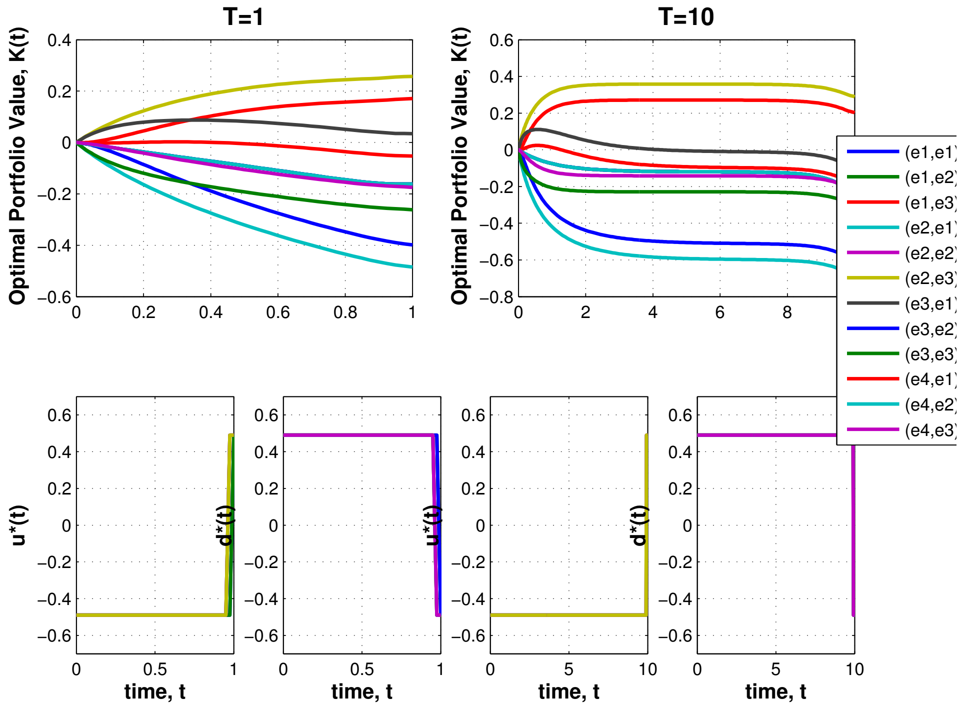


FIG 12. *Solution to problem C.3.3. Minimum Return Function* $k(t, z, x)$*, optimal up controls* $u^*(t, z, x)$ *and down controls* $d^*(t, z, x)$ *for the self-financing portfolio with maximal terminal wealth. Figures(a) and (b) are for* $T = 1$ *and* $T = 10$ *respectively. Various* (z, x) *values are represented by the vectors* (e_i, e_j)*.*

of x to the weights and that of e_1, e_2, e_3, e_4 of z to asset prices, but it is only the value matrix V that appears in any of the solutions and these mappings are immaterial.

C.4.3. *Transaction Costs.* If buying/selling of assets incurs a transaction cost then every weight shift is associated with a cost. This can be modeled in terms of the control costs. We can see that a value of $u = -\frac{1}{2}$ represents the case of a minimal rate of buying the first asset, while $u = \frac{1}{2}$ represents a maximal rate of buying the first asset. Likewise, the values $d = -\frac{1}{2}$ to $d = \frac{1}{2}$ represent the range of the rates of selling the first asset. Hence a reasonable metric for the transaction costs would be $(u + \frac{1}{2})^2 + (d + \frac{1}{2})^2$. For example, a performance measure like $(\alpha > 0)$

$$\eta(u, d) = \mathbb{E} \int_0^T \alpha((u + \frac{1}{2})^2 + (d + \frac{1}{2})^2)dt + \mathbb{E}(U(s(T)))$$

APPENDIX D: SUMMARY OF NOTATIONS AND SYMBOLS

A stochastic basis $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is assumed where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathbb{F}$ a filtration $(\mathcal{F}_t)_{t \in T}$ on this space for a totally ordered index set T ($\subseteq \mathbb{R}^+$in our case). All stochastic processes are assumed to be right continuous and adapted to $\mathbb{F}$.

$\mathbb{F}$	A filtration $(\mathcal{F}_t)_{t \in T}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ where T is a totally ordered index set
$\mathbb{G}^n$	The space of square matrices of dimension n of the form $F_{kl} - F_{ll}$ where F_{ij} is the matrix of all zeros except for one in the $i'th$ row and $j'th$ column
$\mathbb{E}^n$	The space of diagonal $n \times n$ matrices with only 1's or 0's
I_n	$n \times n$ identity matrix, $I_n \in \mathbb{E}^n$
$\mathbb{P}^n$	The space of all stochastic matrices of dimension n
$\{e_i\}_{i=1}^n$	The set of n standard basis vectors in $\mathbb{R}^n$
$\phi(t)$	A real-valued function ϕ on $\mathbb{R}^+ \times \{e_i\}_{i=1}^n$ will be written as the vector $\phi(t) \in \mathbb{R}^n$ as $\phi(t,x) = \phi^T(t)x$ where $x \in \{e_i\}_{i=1}^n$
$\mathbf{\Phi}(t)$	A real-valued function ϕ on $\mathbb{R}^+ \times \{e_i\}_{i=1}^r \times \{e_i\}_{i=1}^n$ is written as the $r \times n$ real matrix $\mathbf{\Phi}(t)$ as $\phi(t,z,x) = z^T \Phi(t) x$ where $z \in \{e_i\}_{i=1}^r$ and $x \in \{e_i\}_{i=1}^n$
$A^T(z)K$	Denotes the matrix whose $j'th$ column is $A(e_j)K^T e_j^T$ which can be more explicitly written as $\sum_z A^T(z)Kzz^T$
$\|M\|$	For a matrix M represents the element-by-element absolute value of a matrix
$M^{\cdot 2}$	For a matrix M represents the element-by-element squared
$\mathbf{e}_r$	The $r-$vector $[1\ 1 ... 1]^T$

ACKNOWLEDGEMENTS

Special thanks to Dr. Roger Brockett of Harvard School of Engineering and Applied Sciences who provided inspiration for the problems discussed in this paper, and to Dr. Andrew JK Phillips at Harvard Medical School for valuable feedback.

REFERENCES

[1] Brockett, R. (2008). Optimal control of observable continuous time markov chains. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on.* 4269–4274.

[2] Geerts, T. (1989). All optimal controls for the singular linear-quadratic problem without stability; a new interpretation of the optimal cost. *Linear Algebra and its Applications 116*, 135 – 181.

[3] Gough, J., Belavkin, V. P., and Smolyanov, O. G. (2005). Hamiltonjacobibellman equations for quantum optimal feedback control. *Journal of Optics B: Quantum and Semiclassical Optics* **7**, 10, S237.

[4] Grasman, J. (1980). On a class of nearly singular optimal control problems. *Stichting Mathematisch Centrum. Toegepaste Wiskunde* TW 196/80 (Jan.), 1–12. http://www.narcis.nl/publication/RecordID/oai%3Acwi.nl%3A7657.

[5] Gudder, S. (2008). Quantum Markov chains. *JOURNAL OF MATHEMATICAL PHYSICS* **49**, 7 (JUL).

[6] Kalenscher, T., Tobler, P. N., Huijbers, W., Daselaar, S. M., and Pennartz, C. (2010). Neural signatures of intransitive preferences. *Frontiers in Human Neuroscience* **4**, 49.

[7] Makowski, M. and Piotrowski, E. W. (2005). Quantum Cat's Dilemma. *eprint arXiv:quant-ph/0510110*.

[8] Noh, E.-J. and Kim, J.-H. (2011). An optimal portfolio model with stochastic volatility and stochastic interest rate. *Journal of Mathematical Analysis and Applications* **375**, 2, 510 – 522.

[9] Piotrowski, E. and Makowski, M. (2005). Cat's dilemma - Transitivity vs. intransitivity. *FLUCTUATION AND NOISE LETTERS* **5**, 1 (MAR), L85–L95.

[10] Plateau, B. and Stewart, W. J. (1997). Stochastic automata networks. In *Computational Probability.* Kluwer Academic Press, 113–152.

[11] Zhou, E., Lin, K., Fu, M., and Marcus, S. (2009). A numerical method for financial decision problems under stochastic volatility. In *Winter Simulation Conference (WSC), Proceedings of the 2009.* 1299 –1310.

[12] Zilli, E. A. and Hasselmo, M. E. (2008). Analyses of markov decision process structure regarding the possible strategic use of interacting memory systems. *Frontiers in Computational Neuroscience* **2**, 6.

31 Oxford Street, Cambridge MA 02138,
E-mail: mgupta@fas.harvard.edu